ROAD AND OFF-ROAD VEHICLE SYSTEM DYNAMICS HANDBOOK



Edited by
Giampiero Mastinu • Manfred Ploechl

CRC Press
Taylor & Francis Group

ROAD AND OFF-ROAD VEHICLE SYSTEM DYNAMICS HANDBOOK

ROAD AND OFF-ROAD VEHICLE SYSTEM DYNAMICS HANDBOOK

Edited by

Giampiero Mastinu
Manfred Ploechl

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

The handbook is not to be used as a manual, that is, as a book containing practical instruction on how to build or modify existing vehicles or their subsystems. The publisher and the editors agree that the handbook has the only aim of explaining to readers both the relevant physical phenomena and how the solution of technical problems can be obtained by means of a scientific approach. Technical applications have then to be realized by experienced engineers only on the basis of national or international standards and/or practical rules that are often omitted in the handbook.

To the fullest extent of the law, neither the publisher nor the authors, contributors, or editors assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence, or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2014 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20130830

International Standard Book Number-13: 978-1-4200-0490-8 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface ix
Contributors xi

Chapter 1 History of Road and Off-Road Vehicle System Dynamics 1
Masao Nagai

Part I Modeling, Analysis, and Optimization in Vehicle System Dynamics

Chapter 2 Vehicle Models and Equations of Motion 13
Werner Schiehlen

Chapter 3 Simulation Algorithms and Software Tools 45
Martin Arnold

Chapter 4 Nonlinear Solid Mechanics with Finite Elements 69
Anna Pandolfi

Chapter 5 Nonlinear Vehicle Dynamics 121
Hans True

Chapter 6 Controls and Identification 159
Stefan Jakubek and Martin Kozek

Chapter 7 Actuators and Sensors 221
Yoshihiro Suda

Chapter 8 Optimization of Ground Vehicle Systems 241
Massimiliano Gobbi and Panos Y. Papalambros

Chapter 9 Fatigue and Structural Durability of Automotive Components 263
Thomas Bruder, Holger Hanselka, Rüdiger Heim, Heinz Kaufmann, Michael Kieninger, Jürgen Nuffer, and Cetin M. Sonsino

Chapter 10 Reliability Assessment of Mechatronic Devices in Vehicles 307
Bernd Bertsche, Jochen Gäng, Holger Hanselka, Soong-Oh Han, Jürgen Nuffer, and Kai Wolf

Part II Vehicle Concepts and Aerodynamics

Chapter 11 Conceptual Design of Road Vehicles Related to Dynamics 329
Giampiero R.M. Mastinu

Chapter 12 Off-Road Vehicles (Wheeled and Tracked) 395
Günter H. Hohl

Chapter 13 Motorcycles and Three-Wheeled Vehicles 477
Robin S. Sharp

Chapter 14 Race Cars: Frame, Suspension, Aerodynamics 497
Andrea Toso

Chapter 15 Race Cars: Braking System 517
Carlo Maria Domenico Cantoni and Giorgio Previati

Chapter 16 Aerodynamics and Vehicle Dynamics 529
Andreas Wagner

Part III Pneumatic Tires and Contact Wheel-Road/Off-Road

Chapter 17 Tire as a Vehicle Component 563
Hans B. Pacejka

Chapter 18 Pneumatic Tire Models: The Detailed Mechanical Approach 625
Michael Gipser

Chapter 19 Pneumatic Tire: Construction and Testing 657
Maurizio Boiocchi and Giuseppe Matrascia

Chapter 20 Mechanics of Off-Road Vehicle–Terrain Interaction: Terramechanics 697
Jo Y. Wong

Part IV Modeling of Vehicle Subsystems

Chapter 21 Suspension Systems 727
Wolfgang Matschinsky

Chapter 22 Active and Semiactive Suspension Systems 769
Davor Hrovat, Eric H. Tseng, Michael Fodor, and Jahan Asgari

Chapter 23 Driveline 797
Andreas Laschet and Ferit Küçükay

Chapter 24 Brake System Dynamics 837
Carlo Maria Domenico Cantoni, Riccardo Cesarini, Giampiero R.M. Mastinu, Giorgio Previati, and Roberto Sicigliano

Chapter 25 Steering System 919
Ichiro Kageyama

Chapter 26 Structural and Dynamic Problems in Car Body Design 943
Giovanni Belingardi and Massimiliano Avalle

Part V Vehicle Dynamics and Active Safety

Chapter 27 Basics of Longitudinal and Lateral Vehicle Dynamics 971
Manfred Ploechl, Peter Lugner, and Johannes Edelmann

Chapter 28 Detailed Modeling, Simulation, and Analysis of Vehicle Dynamics 1023
Dieter Ammon

Chapter 29 Ride Comfort and Road Holding 1059
Karl Popp

Chapter 30 Control of Horizontal Vehicle Motion 1093
Anton van Zanten

Chapter 31 Active and Semi-Active Suspension Control 1179
Davor Hrovat

Chapter 32 Integrated Controls 1227
Masato Abe

Chapter 33 Dynamics of Heavy Commercial Vehicles and Buses 1245
John Aurell

Chapter 34 Dynamics of Off-Road Vehicles 1279
Jo Y. Wong

Chapter 35 Motorcycle Handling Dynamics 1313
Robin S. Sharp

Part VI Man–Vehicle Interaction

Chapter 36 Vehicle Comfort 1379
Karl Siebertz

Chapter 37 Subjective and Objective Evaluations of Car Handling and Ride 1395
Gwanghun Gim

Chapter 38 Driver Models in Automobile Dynamics Application 1477
Manfred Ploechl and Johannes Edelmann

Part VII Intelligent Vehicle Systems

Chapter 39 Automatic Lateral Vehicle Control 1525
Huei Peng

Chapter 40 Longitudinal Control 1549
Paul Fancher and Charles MacAdam

Part VIII Road Accident Reconstruction and Passive Safety

Chapter 41 Analysis and Reconstruction of Road Accidents 1583
Horst Ecker

Chapter 42 Automotive Structural Crashworthiness and Occupant Protection 1611
Jorge Ambrósio

Index 1663

Preface

Some years ago, Swets & Zeitlinger Publishing Company, together with International Association of Vehicle System Dynamics (IAVSD) (represented both by the former Secretary General Professor Robin Sharp and by the former editor in chief of *Vehicle System Dynamics*, Professor Peter Lugner), sent us a proposal to publish a handbook on the dynamics of road and off-road vehicles. Since then, many efforts have been made to gather distinguished contributors to the project. After Swets & Zeitlinger was acquired by Taylor & Francis Group, these efforts were intensified, and we are now finally able to show the result to the scientific and technical community.

Our aim was to publish an authoritative set of standard reference works in the area of vehicle system dynamics. Contemporary publications on the topic are either very specialized (scientific papers) or rather time consuming to read (dedicated books). This handbook is targeted toward both the beginner and the experienced engineer. It can also be used as a textbook for master's courses at universities.

The chapters of the handbook have been written by highly reputed scientists and engineers; many of them have already written full books on the topic they have dealt with in the handbook. The authors have been selected both from industry and universities in order to provide a balance between scientific and rather practical approaches.

The handbook has been written by more than 50 authors from 12 countries in Europe, America, and Asia. It consists of 42 chapters, and 23 universities and 9 prominent companies have been directly or indirectly involved in the project.

The handbook has been conceived as a series of independent chapters covering the state of the art and has been divided into eight parts (Modeling, Analysis, and Optimization in Vehicle System Dynamics; Vehicle Concepts and Aerodynamics; Pneumatic Tires and Contact Wheel-Road/Off-Road; Modeling of Vehicle Subsystems; Vehicle Dynamics and Active Safety; Man–Vehicle Interaction; Intelligent Vehicle Systems; and Road Accident Reconstruction and Passive Safety) and begins with a historical introduction.

We have tried not to overemphasize automobiles and have covered motorcycles, commercial vehicles, and off-road vehicles as well. The main focus, however, is on automobiles.

We have provided the authors with a free hand to express their remarkable experience in their chapters. The contributions were later balanced by a careful reading of all the chapters by the editors. The topics that were not covered have been respectively requested to the proper authors.

We are fully aware that the handbook will not remain up-to-date forever. Even the delay between the contributions written first and the publication date is not negligible. Anyway, the contributions have been written in a plain, informative manner and are centered on basic issues: they will be useful for a relatively long time.

The handbook is a collection of formulae and/or plots that help solve problems without any mathematical proof (left to a subsequent reading of papers or books cited in the literature). Minor technical problems have been omitted, and only high-level scientific/technical problems have been dealt with. We have tried to follow the high standards set by the book on the mechanics of pneumatic tires edited by S. K. Clark, which was published by the U.S. Department of Transportation in 1981.

We wanted this book to mirror the scientific/technical approaches that are used worldwide to solve vehicle system dynamics design problems. We have therefore allowed different views of the same topic given by different authors, assigning responsibility to the reader for a (guided) critical usage of the information collected in the handbook. Engineering is a stimulating, complex field, and allowing a single opinion on a still open question would have considerably reduced the richness of the handbook.

The handbook can be used starting from… the *bottom*, that is, by referring to the subject index to find the desired topic in the book. The experienced reader can go through the book in any order. The inexperienced reader may benefit from studying the "horizontal" chapters, positioned at the beginning of the handbook, that deal with general topics (such as system modeling, analysis, and optimization, Chapters 2 through 10). Such horizontal chapters are the basis for the subsequent reading of the "vertical" chapters that deal with special topics (such as tires and vehicle subsystems, man–vehicle interaction, and active and passive safety, Chapters 11 through 42). In any case, the subject index can help the reader identify the parts of the handbook (belonging either to the horizontal or to the vertical chapters) that are of his or her interest.

As mentioned earlier, the handbook is devoted both to the beginners wishing to acquire a basic knowledge on a topic and to the experienced engineer or scientist wishing to have an up-to-date information on a particular field. This apparently conflicting aim was achieved by asking the authors to deal with basic issues and to add proper references to specialized papers for further reading.

We would like to thank all the authors for their great efforts and for having participated enthusiastically in the project. They have patiently edited and revised their chapters and have provided a final shape to the handbook, reflecting current knowledge on road and off-road vehicle system dynamics.

Giampiero R.M. Mastinu
Milan, Italy

Manfred Ploechl
Vienna, Austria

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508-647-7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com

Contributors

Masato Abe
Department of Vehicle System Engineering
Kanagawa Institute of Technology
Atsugi-shi, Japan

Jorge Ambrósio
Institute of Mechanical Engineering
Instituto Superior Tecnico
Technical University of Lisbon
Lisbon, Portugal

Dieter Ammon
Daimler AG
Stuttgart, Germany

Martin Arnold
Institute of Mathematics
University of Halle-Wittenberg
Halle, Germany

Jahan Asgari
Research and Innovation Center
Ford Motor Company
Dearborn, Michigan

John Aurell
Volvo Trucks
Gothenburg, Sweden

Massimiliano Avalle
Department of Mechanical and Aerospace Engineering
Politecnico di Torino
Turin, Italy

Giovanni Belingardi
Department of Mechanical and Aerospace Engineering
Politecnico di Torino
Turin, Italy

Bernd Bertsche
Institute of Machine Components
University of Stuttgart
Stuttgart, Germany

Maurizio Boiocchi
Pirelli Tire S.p.A.
Milan, Italy

Thomas Bruder
BMW Group
Munich, Germany

Carlo Maria Domenico Cantoni
Brembo S.p.A.
Bergamo, Italy

Riccardo Cesarini
Brembo S.p.A.
Bergamo, Italy

Horst Ecker
Institute of Mechanics and Mechatronics
Engineering Dynamics
Vienna University of Technology
Vienna, Austria

Johannes Edelmann
Institute of Mechanics and Mechatronics
Vehicle System Dynamics Research Group
Vienna University of Technology
Vienna, Austria

Paul Fancher
Transportation Research Institute
The University of Michigan
Ann Arbor, Michigan

Michael Fodor
Research and Innovation Center
Ford Motor Company
Dearborn, Michigan

Jochen Gäng
Institute of Machine Components
University of Stuttgart
Stuttgart, Germany

Gwanghun Gim
Cheng Shin Tire & Rubber Co., Ltd.
Kunshan, Jiangsu, People's Republic of China

Michael Gipser
Department of Automotive Engineering
Esslingen University of Applied Sciences
Esslingen, Germany

Massimiliano Gobbi
Department of Mechanical Engineering
Politecnico di Milano
Milan, Italy

Soong-Oh Han
Freudenberg Forschungsdienste SE & Co. KG
Weinheim, Germany

Holger Hanselka
Fraunhofer Institute for Structural Durability and System Reliability LBF
Darmstadt, Germany

Rüdiger Heim
Fraunhofer Institute for Structural Durability and System Reliability LBF
Darmstadt, Germany

Günter H. Hohl
Austrian Society of Automotive Engineers
Vienna, Austria

Davor Hrovat
Research and Innovation Center
Ford Motor Company
Dearborn, Michigan

Stefan Jakubek
Division of Control and Process Automation
Institute of Mechanics and Mechatronics
Vienna University of Technology
Vienna, Austria

Ichiro Kageyama
Director of Nihon University Center for Automotive Research (NU-CAR)
Department of Mechanical Engineering
College of Industrial Technology
Nihon University, Japan

Heinz Kaufmann
Fraunhofer Institute for Structural Durability and System Reliability LBF
Darmstadt, Germany

Michael Kieninger
Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS
Sankt Augustin, Germany

Martin Kozek
Division of Control and Process Automation
Institute of Mechanics and Mechatronics
Vienna University of Technology
Vienna, Austria

Ferit Küçükay
Institute of Automotive Engineering
Braunschweig University of Technology
Braunschweig, Germany

Andreas Laschet
ARLA Maschinentechnik GmbH
Wipperfürth, Germany

Peter Lugner
Institute of Mechanics and Mechatronics
Vehicle System Dynamics Research Group
Vienna University of Technology
Vienna, Austria

Charles MacAdam
The University of Michigan
Ann Arbor, Michigan

Giampiero R.M. Mastinu
Department of Mechanical Engineering
Politecnico di Milano
Milano, Italy

Giuseppe Matrascia
Pirelli Tire S.p.A.
Milan, Italy

Wolfgang Matschinsky
Büssing AG
Brunswick, Germany

and

BMW AG
Munich, Germany

Masao Nagai
Faculty of Engineering
Department of Mechanical Systems Engineering
Tokyo University of Agriculture and Technology
Tokyo, Japan

Jürgen Nuffer
Fraunhofer Institute for Structural Durability and System Reliability LBF
Darmstadt, Germany

Hans B. Pacejka
Delft University of Technology
Delft, the Netherlands

Anna Pandolfi
Department of Civil and Environmental Engineering
Politecnico di Milano
Milan, Italy

and

Division of Engineering and Applied Sciences
California Institute of Technology
Pasadena, California

Panos Y. Papalambros
Department of Mechanical Engineering
The University of Michigan
Ann Arbor, Michigan

Huei Peng
Department of Mechanical Engineering
The University of Michigan
Ann Arbor, Michigan

Manfred Ploechl
Institute of Mechanics and Mechatronics
Vehicle System Dynamics Research Group
Vienna University of Technology
Vienna, Austria

Karl Popp
Institute of Dynamics and Vibration Research
Leibniz Universitaet Hannover
Hannover, Germany

Giorgio Previati
Department of Mechanical Engineering
Politecnico di Milano
Milano, Italy

Werner Schiehlen
Institute of Engineering and Computational Mechanics
University of Stuttgart
Stuttgart, Germany

Robin S. Sharp
Faculty of Physical and Engineering Sciences
Department of Mechanical Engineering Sciences
University of Surrey
Guildford, United Kingdom

Roberto Sicigliano
Brembo S.p.A.
Bergamo, Italy

Karl Siebertz
Ford Forschungszentrum
Aachen, Germany

Cetin M. Sonsino
Fraunhofer Institute for Structural Durability and System Reliability LBF
Darmstadt, Germany

Yoshihiro Suda
The University of Tokyo
Tokyo, Japan

Andrea Toso
Dallara
Parma, Italy

Hans True
DTU Compute
Technical University of Denmark
Lyngby, Denmark

Eric H. Tseng
Research and Innovation Center
Ford Motor Company
Dearborn, Michigan

Andreas Wagner
Audi AG Ingolstadt
Ingolstadt, Germany

Kai Wolf
Institute for Security Systems
University of Wuppertal
Velbert, Germany

Jo Y. Wong
Department of Mechanical and Aerospace Engineering
Carleton University
Ottawa, Ontario, Canada

Anton van Zanten (retired)
Robert Bosch GmbH
Abstatt, Germany

1 History of Road and Off-Road Vehicle System Dynamics

Masao Nagai

CONTENTS

1.1 Components 1
1.2 Modeling 2
1.2.1 Vehicle Model 2
1.2.2 Tire Model 4
1.2.3 Driver Model 4
1.3 Chassis Control 6
1.4 Intelligent Transport Systems 7
1.5 Analysis Tools 8
1.6 Related Organizations and Events 8
References 9

It has been more than 100 years since the world's first motor vehicle was invented (Figure 1.1). Since then, automobiles have been widely accepted in society with the progress of modern industry. Moreover, as shown in Table 1.1, since the invention of the first motor vehicles, chassis technology as well as components such as the tire, steering, suspension, brake, and power train have been extensively studied and equipped on vehicles to secure their handling, driving performance, stability and also to improve ride comfort. Table 1.1 shows that the fundamental theories of vehicle dynamics including modeling of vehicle motion and tire have been presented since the 1950s. These theories lead to a large growth in the study of vehicle dynamics and great development of various types of chassis control technologies during the last 20 years of the twentieth century. This section will brief on history of vehicle system dynamics based on Table 1.1 and a number of excellent state-of-the-art papers in the last two decades.

In the study of vehicle dynamics, it is common to treat the vehicle as a rigid body, which has six degrees of freedom in motion. To take a glance at the technologies of vehicle dynamics in a systematic approach, Figure 1.2 shows a conceptual diagram of the relationship between each degree of freedom of the vehicle, and how each control input relates to each vehicle motion and vehicle performances [1].

1.1 COMPONENTS

At the top of Table 1.1, the history of components equipped in the motor vehicles is shown. The first row shows the development of the driving system, from engine to motor. Due to today's environmental problems, technology of low-emission vehicles has been developed since the last decade of the twentieth century. The second row shows the development of tires. Recently, a tire that can measure its generating forces is now in development process. In the conventional gasoline vehicle, the trend toward using an electric motor cooperating with the engine is increasing in order to improve traction performance and fuel economy. In addition, as for miniaturization of the motor,

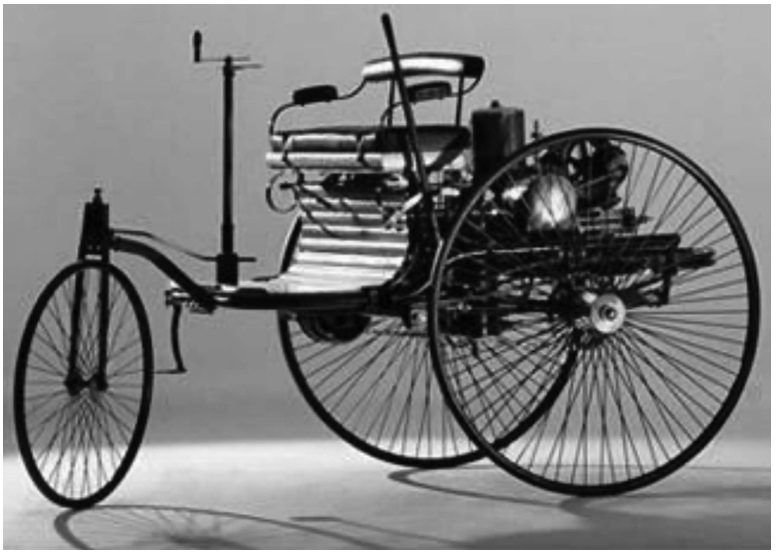


FIGURE 1.1 1886 Benz "Patent Motor Car" first high-speed single-cylinder four-stroke engine.

a sophisticated configuration called the "in-wheel motor" plays an important role as the main driving source of electric vehicle. The third row shows the development of the suspension system, the fourth row shows the development of the steering system, and the last row shows the development of the braking system. As for recent advance in electronic control devices, it is obvious that the mechanical linkage is going to be replaced by "mechatronics."

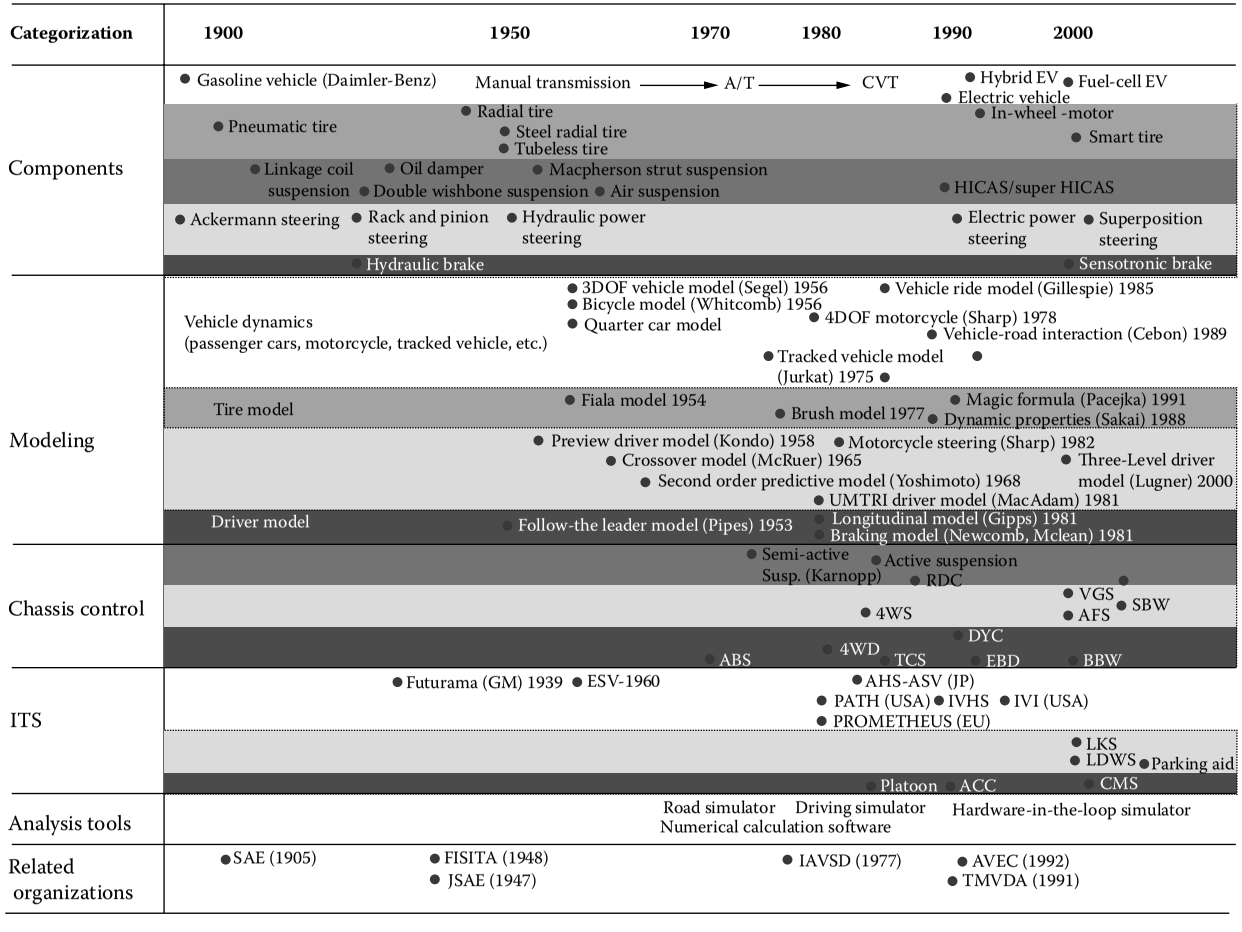
1.2 MODELING

A number of mathematical models used to describe vehicle and tire dynamics have been proposed since the 1950s. From Table 1.1, the modeling of vehicle dynamics can be categorized into three parts: the vehicle model, tire model, and driver model.

1.2.1 Vehicle Model (Chapters 27 through 35)

In modeling vehicle motion, a road vehicle treated as a rigid body has six degrees of freedom, which are lateral, longitudinal, vertical, yawing, pitching, rolling directions. However, in order to analyze characteristics of vehicle motion, a simpler model is more attractive to employ. Segel presented a three-degree-of-freedom model considering lateral, yawing, and rolling motion to describe vehicle-cornering characteristics in 1956 [2]. Moreover, Whitcomb and Milliken presented a well-known two-wheel vehicle model in 1956 neglecting rolling motion and this is a useful model in theoretical analysis of lateral motion and control until now [3]. A quarter car model for analyzing characteristics of a suspension system was also presented around the same time. Gillespie presented a vehicle ride model in 1985 for simulating dynamic tire forces [4]. Besides four-wheel motor vehicles, in 1978, Sharp presented a four-degree-of-freedom model of a motorcycle considering rolling motion in the two-wheel model and also explained the theory of steering behavior of motorcycles by calculating eigenvectors for (straight-line) weave and wobble modes, etc. [5,6]. Modeling of an articulated vehicle is also studied to analyze the braking stability and phenomena of jackknifing. Cebon reviewed vehicle–road interaction characteristics in [7] and contributed a number of technical papers on vehicle-generated road damage and also running on a bridge [8]. Besides models of road vehicles, tracked vehicle dynamics, including mobility over soft terrain, ride dynamics over rough surfaces,

TABLE 1.1
Historic Synopsis of Vehicle System Dynamics



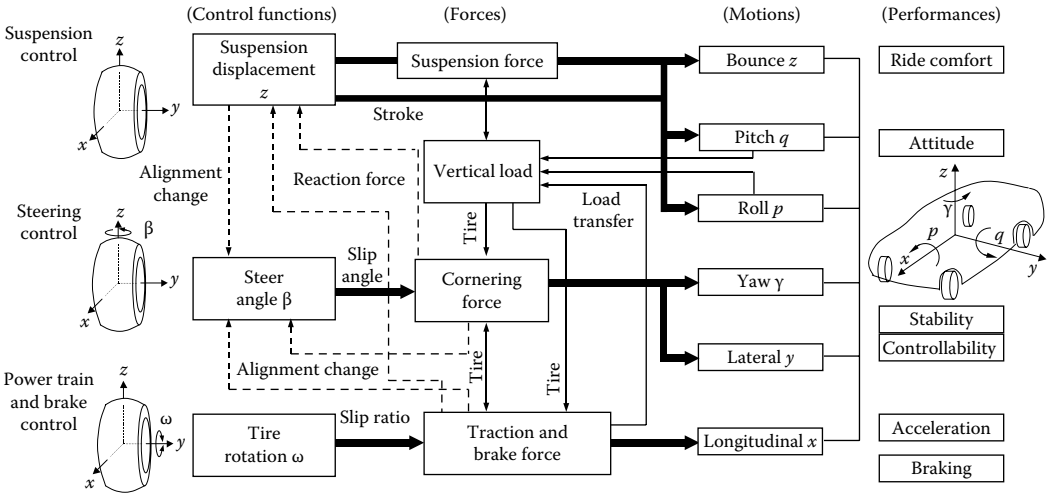


FIGURE 1.2 Conceptual diagram of vehicle dynamics and control. (From Kizu, R. et al., Electronic control of car chassis—Present status and future perspective, *IEEE International Congress on Transportation Electronics*, 88CH2533-8, 1988, pp. 173–188.)

and maneuverability, is also one of our interests. The empirical approach to the modeling of tracked vehicle mobility is best exemplified by the work of the U.S. Army Engineer Waterways Experiment Station in 1975. Wong presented models known as NTVPM for flexible tracks, and RTVPM for rigid link tracks with a long track pitch used in agricultural vehicles and construction vehicles, and reviewed the simulation models [9] (see Chapters 20 and 34).

1.2.2 Tire Model (Chapters 17 through 20)

Tire models play an important key role in chassis control system design and analysis of vehicle dynamics. A very simple approach for the tire model is a spring constant for vertical deflections (as used with quarter car models) and a linear relationship between slip angle and lateral force generation (as used with two-wheel models). Theoretical approaches for modeling nonlinear characteristics of longitudinal and lateral forces of tires with respect to slip phenomenon were presented by Fiala and the well-known Brush model. Mixed longitudinal and lateral slip can be taken into consideration by a characteristic curve, groups of curves, or by suitable approximation functions, for example, works by Bakker et al. [10]. Moreover, the dynamic tire characteristics require further details: in the simplest approaches, a first-order differential equation approximates the tire force buildup, presented by Sakai [11]. For more precise reproduction of forces due to short-wave road unevenness, further refinements of the tire model structure are necessary as found in F-Tire or RmodK, presented by Gipser [12]. A detailed nonlinear dynamic finite element method (FEM) like DNS-Tire, also presented by Gipser, may be necessary in the case of investigations of extreme situations. FEM plays a significant role in the field of tire mechanics as shown by Knothe et al. who reviewed tire–road contact mechanics in [13]. Moreover, tire–road contact is an important topic in studying vehicle dynamics and stability, as demonstrated by Knothe and Boehm who reviewed running stability and its historic synopsis in [14].

1.2.3 Driver Model (Chapter 38)

Due to the limited scope of this section, the emphasis here is on the control aspect of the human driver and his/her modeling. In the investigation of vehicle dynamics and design of chassis control, human factors are inevitably involved in the procedures. Even though development in chassis design and control can greatly enhance vehicle performance, one thing that cannot be neglected

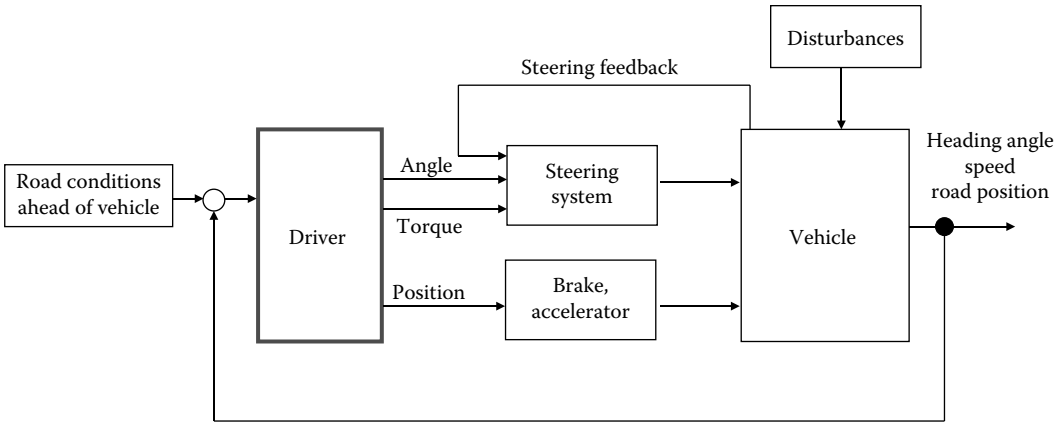


FIGURE 1.3 Hypothetical vehicle control loop including driver.

is the characteristics of the driver who operates the vehicle by recognizing the road environment as a closed-loop system, as shown in Figure 1.3. In addition, in the development of autonomous vehicles and driver assistance systems, human factors become more and more important in the design of such systems than in the past. In near future, the importance of studying the driver model will increase in supporting the design of a driver–environment-adaptive vehicle control system, as shown in Figure 1.4 [15]. From 1950s, the proposed driver modeling can be classified into two groups: the longitudinal driver model and the lateral driver model [16,17]. Complete treatments of the driver modeling effort and associated measurements of driver–vehicle system responses have emerged, particularly with the increased use of computers and driving simulators.

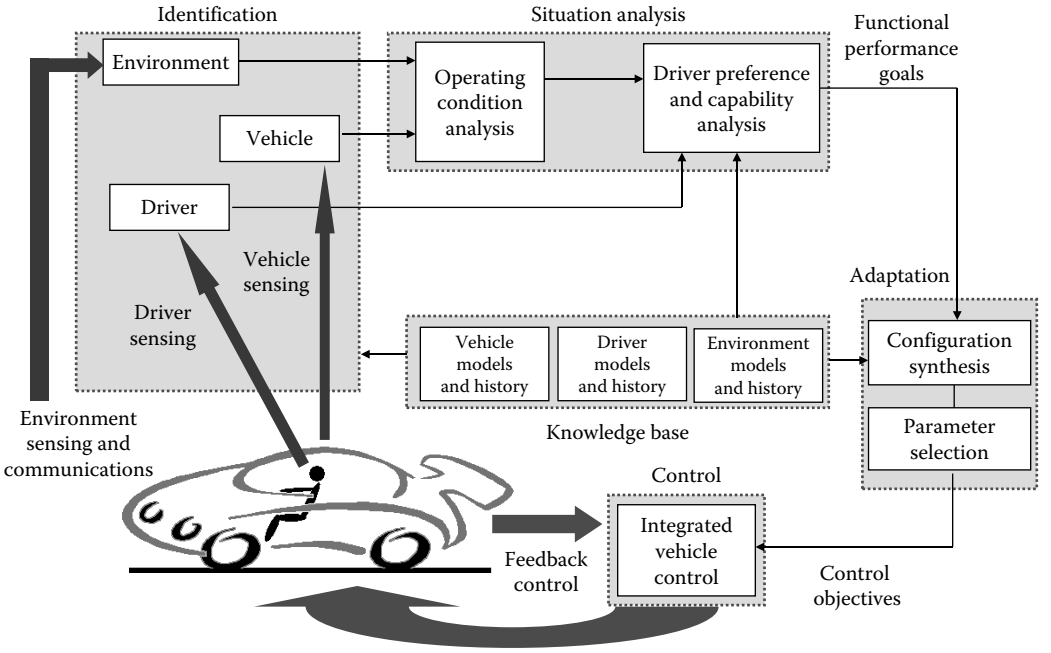


FIGURE 1.4 Design concept of driver–environment–adaptive vehicle control system. (From Schilke, N.A. et al. Integrated vehicle control, *Proceedings of International Congress on Transportation Electronics, IEEE/SAE Convergence*, 88, Michigan, USA, 1988, pp. 97–106.)

The concept that the human driver must look ahead becomes the first representative concept in modeling the lateral driver model. Kondo and Ajimine's work focused on measuring the driver "sight point" ahead of the vehicle and developed steering based on deviations between the sight point and the desired course [18]. Yoshimoto included a "second prediction term" into the preview driver model to represent driver anticipation when entering a curve [19]. To model the adaptive control behavior of the human driver, McRuer and Jex offered the now well-known crossover model for automobile drivers [20]. Ploechl and Lugner proposed the three-level driver model, anticipatory control, compensatory control, and local control, and explained its application to driving simulations [21]. There is also an attempt to develop a driver model capability applicable to nonlinear, near-limit operating conditions, for example, when a tire blows out, as in the GM/UMTRI (General Motors/The University of Michigan, Transportation Research Institute) driver model presented by MacAdam [22].

Examination of technical papers related to the longitudinal driver model reveals that models and associated studies of driver braking/acceleration control behavior are relatively sparse when compared to work related to lateral steering control. Pipes suggested a linear follow-the-leader model, in which the driver accelerates the vehicle in proportion to the speed difference between the leading vehicle and the following vehicle, in 1953 [23]. Gipps proposed a switching vehicle speed model based on two competing considerations: to keep a safe distance from the lead vehicle and to converge to the desired free flow speed [24]. Newcomb and McLean provide examples of modeling of the driver–vehicle system during longitudinal braking [25]. As for the development of adaptive cruise control (ACC), researchers employ a longitudinal vehicle model to design and analyze the behavior. The works of Fancher et al. are primarily concerned with the headway control behavior of typical drivers and their interaction with ACC systems during highway driving [26]. Peng's recent work using a modified version of Gipps' model is likewise advancing the modeling effort in the headway control area and accurately represents traffic flow behavior of human drivers [27].

1.3 CHASSIS CONTROL [28,29] (CHAPTERS 30 THROUGH 32)

In the last three decades, in the field of vehicle system dynamics, the largest number of technical papers presented and published is on the topic of chassis control system development, in both theoretical and experimental studies. As shown in Figure 1.2, the vehicle chassis control is, in general, to control lateral, vertical, and longitudinal vehicle motions in order to improve handling performance, ride comfort, traction/braking performance, and active safety of the vehicle. As for the major developments of vehicle modeling including tire modeling, the theory of vehicle dynamics shows its potential in applications to the design of chassis control technology to enhance driving performance as well as driving safety. The first representative chassis control technology is the brake system, which is well known as the "antilock-braking system" (ABS). At present, it is the standard equipment in passenger cars. To enhance ride comfort of the vehicle, employing a quarter car vehicle model aids in the design of semi-active suspension and active suspension. Active suspension control can also be adapted to the design of roll distribution control to improve roll stability, especially in heavy duty vehicles. To improve traction performance especially on slippery roads, a traction control system (TCS) has been presented and now is a standard equipment in high-class passenger cars along with ABS.

From the mid-1980s, researchers paid more attention to the chassis control technology in lateral motion. As vehicle speed increases, the vehicle tends to lose its cornering stability due to the increment of centrifugal acceleration. The first technology that improved vehicle lateral motion is four-wheel steering system (4WS). Steering the rear wheel together with the front wheel improves not only low-speed maneuverability but also high-speed stability and handling response. Most of the research on 4WS uses zero-sideslip response as a control objective; thus, the ratio of rear steer angle to front steer angle becomes a function of vehicle speed. In real applications, Sano et al. proposed a type of 4WS in which the rear wheel is steered as a nonlinear function of steering wheel angle [30].

Sato et al. proposed a type of 4WS, which makes use of feedforward of steering wheel angle and feedback of yaw rate to make the control system more robust to external disturbances [31]. Shibahata et al. proposed a 4WS system, which uses a feedforward control for controlling rear wheel steering angle dynamically [32]. A feature of this system is that the rear wheels are steered in the opposite direction to the front wheels when the steering wheel is turned quickly. However, since 4WS control depends on the tire lateral force, in a range where the lateral acceleration is comparatively high, the lateral force becomes saturated and the 4WS cannot perform its control effectively anymore.

To overcome this problem, a new strategy using tire longitudinal forces to control vehicle lateral motion was proposed, and called direct yaw-moment control (DYC). At the International Conference on Advanced Vehicle Control, AVEC'92, there were a number of technical papers on this technology. As a common approach for vehicles in productions, by utilizing the existing ABS hardware or traction force distribution control, DYC system can be realized in real applications. As a representative concept of using DYC to stabilize vehicle motion, Shibahata et al. proposed a "β-method" which points out that instability of vehicle lateral motion is caused by the decrease of restoring yaw moment as the vehicle-sideslip angle becomes larger [33]. It is suggested that DYC should be designed not only to control the vehicle yaw motion but also to control the sideslip motion. Nowadays, in the automobile market, DYC is commonly used for the objective of sideslip prevention, and also for rollover prevention, especially for commercial vehicles. However, control performance of DYC greatly depends on sideslip angle sensing and estimation accuracy, which becomes one of the most important topics in recent years. Since the sideslip angle within the nonlinear region is closely related to the friction coefficient μ between the tire and road surface, it is important to simultaneously detect or estimate the μ value. A number of studies on estimation of β and μ and the sensing and estimation of tire lateral force have been reported since the sessions of Sensors and Actuators in AVEC'94.

In the last few years of the twentieth century, researchers paid attention to the active steering control of the front wheel by electronic control instead of conventional mechanical linkage. The technique of steer-by-wire (SBW) system or power-assisted steering (HPS, EPS) is applied to the steering system to realize active front steering (AFS) [34]. Additional steering control of the front wheel is proposed, to make steering maneuvers easier for drivers, such as in variable gear-ratio steering, and for disturbance compensation, for example, μ-split, wind gust. To improve the driver's feel for steering, techniques of steering reaction torque control are also applied to the system. By analyzing the control characteristics of the driver–vehicle closed-loop system, the design of active control of the front wheel angle and steering reaction torque can be considered in various approaches.

Moreover, combining these chassis control technologies can more effectively enhance vehicle handling and stability. There are many approaches of integrated control, which have been proposed by many researchers. On lateral control, technique of a model following control method is applied to determine each control input to track the desired steering responses [35–37]. In addition, study on optimizing tire force usage for vehicle dynamics control is being conducted in order to achieve maximum vehicle dynamics performance from the viewpoint of tire limit performance [38,39]. Recently, Gordon et al. also provided the scope, methodologies, and architectures for the design and development of interacting control systems in road vehicles in [40].

1.4 INTELLIGENT TRANSPORT SYSTEMS [41–44] (CHAPTERS 39 AND 40)

Following the contents of this handbook, the history of intelligent vehicle–highway systems (IVHS) and advanced vehicle control systems (AVCS) will be mentioned here. The control systems can be categorized into lateral control and longitudinal control. Research on lateral control mainly focuses on the use of an automatic steering device to track lanes automatically. The lane sensing system can be considered as a structure using inductive cables and magnetic markers embedded in the pavement along the lane centerline, or as a machine vision system, with cameras mounted on the vehicle, and image processing software used to identify the location and alignment of the vehicle relative to

the lane markers. Machine vision for road vehicle guidance is recently becoming the main approach to real systems. Applications of this system are the lane-keeping assistance system and the lane departure avoidance system. Besides the lane-keeping control system which focuses mainly in high-speed regions, a driver assistance system in low-speed regions like the Parking Assistance System utilizing an electric power steering device to park a car in the position that the driver desires can also be categorized in this field.

For almost four decades, vehicle longitudinal control has a longer and richer history relevant to AVCS than lateral control. While lateral control is a function of each individual vehicle relative to the roadway reference, longitudinal control adds the complication of potential interactions among multiple controlled vehicles. Control algorithms proposed can be roughly grouped into two categories: (1) fixed block control or point following control and (2) vehicle following control. The largest volume of research in this area is in vehicle-following longitudinal control systems, with the objectives of providing greater lane capacity and reducing traffic congestions. The works on ACC and platoon driving are the main core of this field. An interesting aspect of longitudinal control for automated vehicles is the concept of "string stability." The term string stability refers to the propagation of a vehicle speed perturbation, in other words the amplified response of distance between vehicles, from one end toward the other end of a string of vehicles. For longitudinal collision avoidance, there are systems like head-on crash warning and control systems utilizing automatic braking system such as the collision mitigation system (CMS), which has been already released in the Japanese market.

1.5 ANALYSIS TOOLS [45] (CHAPTERS 2 THROUGH 10)

Numerical simulations are increasingly focused on ride and handling development processes. The simplest models like mass point models and two wheel models can be expressed in classical programming languages such as FORTRAN, C, and C++. They can also be applied with system simulation software such as MATLAB®/Simulink® and Matrix/System Build. Complete solutions for calculating complex models are available on the market, such as Ve-Dyna or Carsim that shows attractive animation. Multi-Body System (MBS) tools like ADAMS from MDI, DADS from LMS, and Simpack from INTEC are based on the simulation of mechanical basic elements such as mass, joints, and springs, and are complemented by numerous component models.

Another way to investigate the vehicle dynamics or effectiveness of the proposed chassis control system by avoiding proving ground tests is using a simulator. Hardware-in-the-Loop (HIL) Simulation techniques allow the integration of real components, for example, tire, steering, braking, etc., into the loop of numerical calculations, which is useful in evaluating and validating the component on virtual driving environments. The driving simulator, specifically the human driver-in-the-loop simulator, is well known for studying the human–machine interface, vehicle handling investigations, and ride comfort.

1.6 RELATED ORGANIZATIONS AND EVENTS

Finally, in reviewing the history of vehicle dynamics, it is impossible to neglect the associations and symposiums that have been organized with the purpose of bringing engineers and researchers from around the world together to share ideas and advance the technological development of the automobile. Society of Automotive Engineers (SAE), founded in 1905, was the first organization deeply related to motor vehicle technology. Next, the International Federation of Automotive Engineer Societies (FISITA) was found in 1948 in Paris and organizes world congress every 2 years. The International Association of Vehicle System Dynamics (IAVSD), found in 1977 in Vienna, is the organization most related to the topics on this handbook. IAVSD organizes a symposium every 2 years where technical papers on not only road vehicles but also rail vehicles are presented and discussed. The International Symposium on AVEC was first organized in 1992 in Yokohama, Japan,

in order to share information on control technology related to automobile applications. At last, the International Tyre Colloquium: Tyre Models for Vehicle Dynamics Analysis (TMVDA), first held in 1991 in the Netherlands, has been organized to address recent investigations and developments in the field of road–tire–vehicle interactions.

Finally, from the reviewed history, we hope the reader could gain a macroscopic view of vehicle system dynamics, and realize that there still remains a number of research challenges and issues to analyze vehicle dynamics and implement vehicle advanced control technologies for future vehicles. The contents that follow this chapter will be extensively contributed in detail by experts in their specialized fields.

REFERENCES

1. Kizu, R. et al., Electronic control of car chassis—Present status and future perspective, *IEEE International Congress on Transportation Electronics*, 88CH2533-8, 1988, pp. 173–188.
2. Segel, L., Theoretical prediction and experimental substantiation of the response of the automobile to steering control, *Proceedings of the Automobile Division*, The Institution of Mechanical Engineers, London, U.K., 1956, pp. 310–330.
3. Whitcomb, D.W. and Milliken, W.F., Design implications of a general theory of automobile stability and control, *Proceedings of the Automobile Division*, The Institution of Mechanical Engineers, London, U.K., 1956, pp. 83–107.
4. Gillespie, T.D., *Fundamental of Vehicle Dynamics*, Society of Automotive Engineers, Warrendale, PA, 1992.
5. Sharp, R.S., The lateral dynamics of motorcycles and bicycles, *Vehicle System Dynamics*, 14(4–6), 1985, 265–284.
6. Sharp, R.S., Stability control and steering responses of motorcycles, *Vehicle System Dynamics*, 35(4–5), 2001, 291–318.
7. Cebon, D., *Handbook of Vehicle–Road Interaction*, Swets & Zeitlinger Publishers, Lisse, the Netherlands, 1999.
8. Cebon, D., Vehicle-generated road damage: A review, *Vehicle System Dynamics*, 18(1–3), 1989, 107–150.
9. Wong, J.Y., Dynamics of tracked vehicles, *Vehicle System Dynamics*, 28(2–3), 1997, 197–220.
10. Bakker, E., Nyborg, L., and Pajecka, H.B., Tyre modelling for use in vehicle dynamics studies, SAE Paper 870421, 1987.
11. Sakai, H., Theoretical and experimental studies on the dynamical properties of tyres. Part 1: Review of theories of rubber friction, *International Journal of Vehicle Design*, 2(1), 1981, 78–110.
12. Gipser, M., Reifenmodelle in der Fahrdynamik, *Proceedings of MKS-Simulation in der Automobilindustrie*, SFT Graz, Austria, 2001.
13. Knothe, K. et al., Advanced contact mechanics—Road and rail, *Vehicle System Dynamics*, 35(4–5), 2001, 361–407.
14. Knothe, K. and Boehm, F., History of stability of railway and road vehicles, *Vehicle System Dynamics*, 31(5–6), 1999, 283–324.
15. Schilke, N.A. et al. Integrated vehicle control, *Proceedings of International Congress on Transportation Electronics, IEEE/SAE Convergence*, Michigan, 88, USA, 1988, 97–106.
16. Guo, K. et al. Modelling of driver/vehicle directional control system, *Vehicle System Dynamics*, 22(3–4), 1993, 141–184.
17. MacAdam, C.C., Understanding and modeling the human driver, *Vehicle System Dynamics*, 40(1–3), 2003, 101–132.
18. Kondo, M. and Ajimine, A., Driver's sight point and dynamics of the driver-vehicle system related to it, *Proceedings of the SAE Automotive Engineering Congress*, Detroit, MI, 1968.
19. Yoshimoto, K., Simulation of man-automobile systems by the driver's steering model with predictability, *Bulletin of the Japan Society of Mechanical Engineers*, 12(51), 1969, 495–500.
20. McRuer, D.T. and Jex, H.R., A review of quasi-linear pilot models, *IEEE Transactions on Human Factors in Electronics*, HFE-8, 1967, 231–249.
21. Ploechl, M. and Lugner, P., A 3-level driver model and its application to driving simulations, *Vehicle System Dynamics*, 33(Suppl.), 2000, 71–82.
22. MacAdam, C.C., Development of a driver model for near/at-limit vehicle handling, UMTRI-2001-43, 2001.

23. Pipes, L.A., An operation analysis of traffic dynamics, *Journal of Applied Physics*, 24, 1953, 271–281.
24. Gipps, P.G., A behavioral car-following model for computer simulation, *Transportation Research B*, 15B, 1981, 105–111.
25. McLean, D., Newcomb, T.P., and Spurr, R.T., Simulation of driver behavior during braking, *Proceedings of the International Mechanical Engineers Conference on Braking of Road Vehicles*, Loughborough, U.K., 1976.
26. Fancher, P., Bareket, Z., and Ervin, R., Human-centered design of an ACC-with-braking and forward-crash-warning system, *Vehicle System Dynamics*, 36(2–3), September 2001, 203–224.
27. Peng, H., Evaluation of driver assistance systems—A human centered approach, *Proceedings of the 6th International Symposium on Advanced Vehicle Control*, Hiroshima, Japan, 2002, pp. 17–24.
28. Furukawa, Y., Yuhara, N., Sano, S., Takeda, H., and Matsushita, Y., A review of four-wheel steering studies from the viewpoint of vehicle dynamics and control, *Vehicle System Dynamics*, 18(1–3), 1989, 151–186.
29. Furukawa, Y. and Abe, M., Advanced chassis control systems for vehicle handling and active safety, *Vehicle System Dynamics*, 28(2–3), 1997, 59–86.
30. Sano, S., Furukawa, Y., and Shiraishi, S., Four wheel steering system with rear wheel steer angle controlled as a function of steering wheel angle, 1986, SAE Paper 860625.
31. Sato, H., Hirota, A., and Yanagisawa, T., Dynamic characteristics of a whole wheel steering vehicle with yaw velocity feedback rear wheel steering, *Proceedings of IMechE*, 1983, C124/83.
32. Shibahata, Y., Irie, N., Itoh, H., and Nakamura, K., The development of an experimental four-wheel-steering system, 1986, SAE Paper 860623.
33. Shibahata, Y., Shimada, K., and Tomari, T., The improvement of vehicle maneuverability by direct yaw moment control, *Proceedings of the 1st International Symposium on Advanced Vehicle Control*, Yokohama, Japan, 1992, pp. 452–457.
34. Segawa, M., Nishizaki, K., and Nakano, S., A study of vehicle stability control by steer by wire system, *Proceedings of the 5th International Symposium on Advanced Vehicle Control*, Michigan, USA, 2000, pp. 233–240.
35. Shino, M., Raksincharoensak, P., and Nagai, M., Vehicle handling and stability control by integrated control of direct yaw moment and active steering, *Proceedings of the 6th International Symposium on Advanced Vehicle Control*, Hiroshima, Japan, 2002, pp. 25–32.
36. Mokhiamar, O. and Abe, M., Combined lateral force and yaw moment control to maximize stability as well as vehicle responsiveness during evasive maneuvering for active vehicle handling safety, *Vehicle System Dynamics*, 37(Suppl.), 2002, 246–256.
37. Heinzl, P., Lugner, P., and Ploechl, M., Stability control of a passenger car by combined additional steering and unilateral braking, *Vehicle System Dynamics*, 37(Suppl.), 2002, 221–233.
38. Mokhiamar, M. and Abe, M., Simultaneous optimal distribution of lateral and longitudinal tire forces for the model following control, *Transactions of the ASME, Journal of Dynamic Systems, Measurement and Control*, 126(4), December 2004, 753–763.
39. Ono, E., Hattori, Y., Muragishi, Y., and Koibuchi, K., Vehicle dynamics integrated control for four-wheel-distributed steering and four-wheel-distributed traction/braking systems, *Vehicle System Dynamics*, 44(2), February 2006, 139–152.
40. Gordon, T. et al., Integrated control methodologies of road vehicles, *Vehicle System Dynamics*, 40(1–3), 2003, 157–190.
41. Tomizuka, M. and Hedrick, J.K., Advanced control methods for automotive applications, *Vehicle System Dynamics*, 24(6–7), 1995, 449–468.
42. Shladover, S.E., Review of the state of development of advanced vehicle control systems (AVCS), *Vehicle System Dynamics*, 24(6–7), 1995, 551–595.
43. Dickmanns, E.D., Computer vision and highway automation, *Vehicle System Dynamics*, 31(5–6), 1999, 325–344.
44. Palkovics, L. and Fries, A., Intelligent electronic systems in commercial vehicles for enhanced traffic safety, *Vehicle System Dynamics*, 35(4–5), 2001, 227–290.
45. Rauh, J., Virtual development of ride and handling characteristics of advanced passenger cars, *Vehicle System Dynamics*, 40(1–3), 2003, 135–156.

Part I

Modeling, Analysis, and Optimization in Vehicle System Dynamics

2 Vehicle Models and Equations of Motion

Werner Schiehlen

CONTENTS

2.1 Introduction 13
2.2 Kinematics 14
2.2.1 Frames of Reference for Vehicle Kinematics 14
2.2.2 Kinematics of a Rigid Body in an Inertial Frame 16
2.2.3 Kinematics of a Rigid Body in a Moving Reference Frame 22
2.2.4 Kinematics of Multi-Body Systems 23
2.3 Dynamics 25
2.3.1 Inertia Properties 25
2.3.2 Newton–Euler Equations 26
2.3.3 Principles of d'Alembert and Jourdain 28
2.3.4 Energy Considerations and Lagrange's Equations 30
2.4 Equations of Motion 32
2.5 Formalisms for Multi-Body Systems 35
2.5.1 Non-Recursive Formalisms 35
2.5.2 Recursive Formalisms 39
2.5.2.1 Kinematics 39
2.5.2.2 Newton–Euler Equations 41
2.5.2.3 Equations of Motion 41
2.5.2.4 Recursion 41
References 42

2.1 INTRODUCTION

Performance, safety, and comfort of a vehicle are related to its low-frequency motions. The corresponding mechanical models are characterized for all kinds of vehicles by stiff parts represented as rigid bodies and soft components like springs, dampers, and actuators (Figure 2.1). The method of multi-body systems is most appropriate for the analysis of vehicle motions and vibrations up to 50 Hz. In this contribution, the derivation of the equations of motions of multi-body systems is shown step by step up to the computer-aided evaluation of these equations.

The roots and perspectives of multi-body system dynamics have been summarized in Ref. [1]. The fundamentals of vehicle system dynamics are presented in textbooks by Popp and Schiehlen [2], Ammon [3], Dukkipati [4], and Wong [5]. Many references are also available from the proceedings of the biennial international symposia on the dynamics of vehicles on roads and tracks, edited, for example, by Berg and Stensson Trigell [6]. Some computational aspects and characteristic benchmarks of multi-body dynamics are presented in the *Multibody System Handbook* [7], and multi-body computer codes in vehicle system dynamics are considered by Kortüm and Sharp [8].

Starting with kinematics for rigid body vehicle systems, the foundations of dynamics together with the principles of d'Alembert and Jourdain are used to get the equations of motion. Then, some aspects

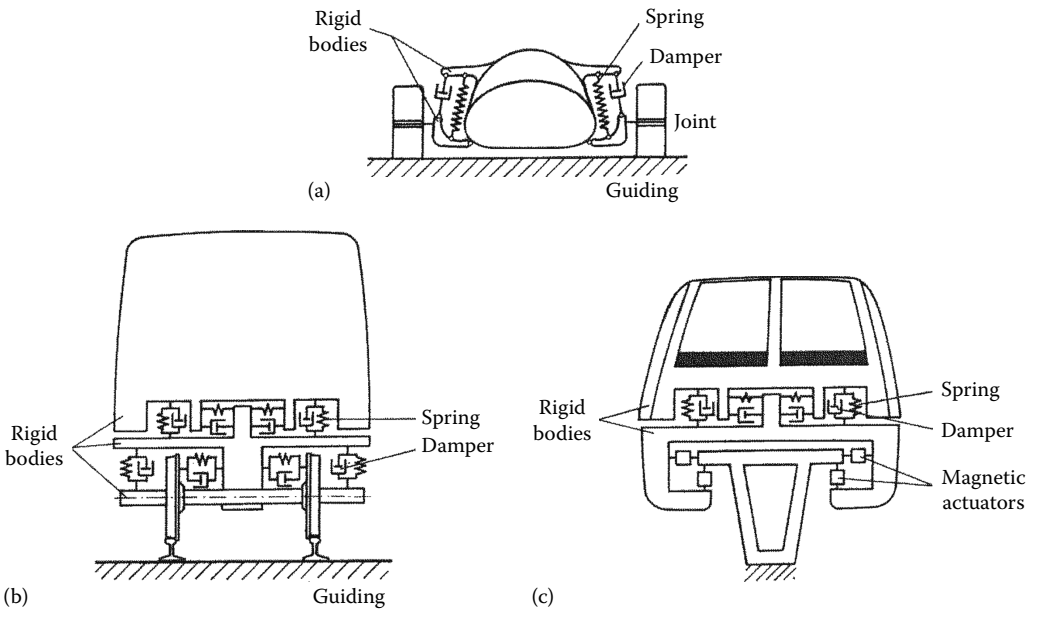


FIGURE 2.1 Multi-body models of vehicles: (a) road vehicle, (b) rail vehicle, and (c) magnetically levitated vehicle.

of multi-body dynamics formalisms and computer codes for vehicle dynamics will be discussed. This contribution is based on original German edition of the monograph by Popp and Schiehlen [2]. Some revisions are found in the English edition and the Chinese translation.

2.2 KINEMATICS

The elements of multi-body systems for vehicle modeling (see Figure 2.1) include rigid bodies, which may also degenerate to particles, coupling elements like springs, dampers, or force-controlled actuators as well as ideal, that is, rigid kinematical connecting elements like joints, bearings, rails, and motion-controlled actuators. The coupling and connection elements are generating internal forces and torques between the bodies of the system and external forces with respect to the environment. Both of them are considered as massless elements.

The kinematical constraints resulting from the connecting elements may be holonomic or non-holonomic, scleronomic or rheonomic, respectively. Holonomic constraints reduce the motion space of the system while non-holonomic constraints reduce the velocity space in addition. The constraint equations are called rheonomic if they depend explicitly on time, or scleronomic otherwise. Real vehicle systems are subject to holonomic constraints only which may be given by geometrical or integrable kinematical conditions. However, in more simplified models, for example, rolling of a rigid wheel or wheelset on a rigid plane, non-holonomic constraints may occur. Some designs of holonomic connecting elements are listed in Table 2.1 depending on the number of degrees of freedom characterizing the remaining possibilities of motion.

Now the motion of vehicle parts will be described mathematically depending on space and time. This is the task of kinematics.

2.2.1 Frames of Reference for Vehicle Kinematics

A prerequisite for the mathematical description of position, velocity, and acceleration of a mechanical system is the definition of appropriate frames of reference. The frames required in vehicle dynamics are shown in Figure 2.2 with the details summarized in Table 2.2. They will be used

TABLE 2.1
Designs for Holonomic Connecting Elements

	Degrees of Freedom		
Motion	**1**	**2**	**3**
Rotary	Revolute joint	Universal joint	Spherical joint
Linear	Prismatic joint	Planar joint	
Mixed	Screw joint	Cylindric joint	General planar joint

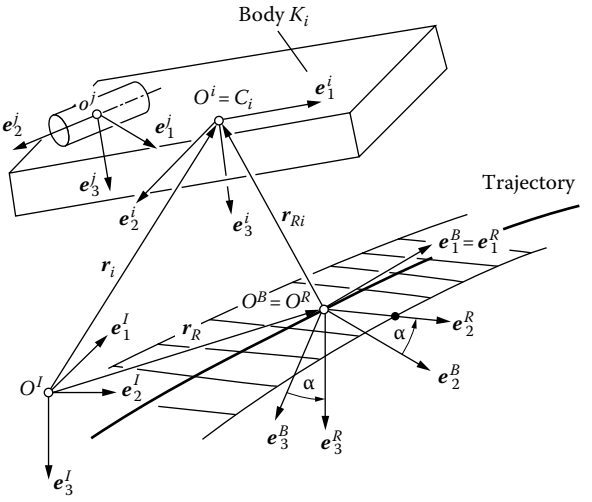


FIGURE 2.2 Frames of reference.

TABLE 2.2
Frames of Reference

Frame of Reference	**Origin of Frame**	**Orientation of Axes**
Inertial frame	O^I	$\boldsymbol{e}_1^I$, $\boldsymbol{e}_2^I$ in horizontal plane
$\left\{O^I, \boldsymbol{e}_\upsilon^I\right\}$	Space fixed	$\boldsymbol{e}_3^I$ perpendicular
Moving trihedron	O^B	$\boldsymbol{e}_1^B \equiv \boldsymbol{e}_t$ tangential to trajectory
$\left\{O^B, \boldsymbol{e}_\upsilon^B\right\}$	Trajectory fixed	$\boldsymbol{e}_2^B \equiv \boldsymbol{e}_1^B$ normal to trajectory
		$\boldsymbol{e}_{31}^B \equiv \boldsymbol{e}_b$ binormal to trajectory
Reference frame	O^R	$\boldsymbol{e}_1^R \equiv \boldsymbol{e}_1^B$
$\left\{O^R, \boldsymbol{e}_\upsilon^R\right\}$	Trajectory fixed	$\boldsymbol{e}_2^R$ in guideway plane
		$\boldsymbol{e}_3^R$ normal to guideway plane
Body-fixed frame	$O^i = C_i$	
$\left\{O^i, \boldsymbol{e}_\upsilon^i\right\}$	Body fixed in center of mass	$\boldsymbol{e}_\upsilon^i$ principle inertia axes
Local frame	O^j	
$\left\{O^j, \boldsymbol{e}_\upsilon^j\right\}$	Arbitrary	$\boldsymbol{e}_\upsilon^j$ locally specified axes

only right-handed Cartesian frames with the unit base vectors $\boldsymbol{e}_\upsilon$, $|\boldsymbol{e}_\upsilon| = 1$ where the Greek indices generally take the integers 1, 2, 3. A basis or frame $\{O, \boldsymbol{e}_\upsilon\}$, respectively, is completely defined by its origin O and its base vectors $\boldsymbol{e}_\upsilon$. For distinction between different frames, the upper right index is used if necessary.

The inertial frame $\left\{O^I, \boldsymbol{e}_\upsilon^I\right\}$ serves as the general reference frame, in particular for the evaluation of the acceleration. The given trajectory of the vehicle is assumed to be a space curve with the moving frame $\left\{O^B, \boldsymbol{e}_\upsilon^B\right\}$ also known as Frenet frame or moving trihedron. The origin O^B is moving with a given speed tangential to the trajectory. The reference frame $\left\{O^R, \boldsymbol{e}_\upsilon^R\right\}$ is closely related to the moving frame. Its origin and the first unit vector coincide with the moving frame $O^R = O^B$, $\boldsymbol{e}_1^R = \boldsymbol{e}_1^B$. The second base vector $\boldsymbol{e}_2^R$, however, is parallel to the guideway surface considering the bank of the road or the track, respectively, pointing to the right with respect to the direction of motion.

The body-fixed frame $\left\{O^i, \boldsymbol{e}_\upsilon^i\right\}$ is the principal axis frame of the rigid body K_i located in its center of mass C_i. This frame describes uniquely the position in space of the body. Finally, there is a defined local frame $\left\{O^j, \boldsymbol{e}_\upsilon^j\right\}$ to describe constraint elements between bodies. It is oriented according to the local specifications like the direction of a joint axis. In the following, a frame is simply identified by its name (upper right index) only.

2.2.2 Kinematics of a Rigid Body in an Inertial Frame

First of all, some definitions and remarks on the nomenclature are presented.

The position of a particle P in space is uniquely defined by the position vector $\boldsymbol{x}$ represented in the inertial frame $\left\{O^I, \boldsymbol{e}_\upsilon^I\right\}$ by its coordinates x_υ as

$$\boldsymbol{x} = x_1\, \boldsymbol{e}_1 + x_2\, \boldsymbol{e}_2 + x_3\, \boldsymbol{e}_3. \tag{2.1}$$

This set of coordinates may be summarized in a column matrix $\boldsymbol{x}^I$ often simply called a vector, that is,

$$\mathbf{x}^{\mathrm{I}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \equiv \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^{\mathrm{T}}, \tag{2.2}$$

where upper right index defines the frame in which the coordinates are measured. This index will often be deleted if there is not any possibility for a mix-up of frames or if there is used only one frame identified in the text. The goal is to present all vector and tensor quantities in one common frame, for example, the inertial frame I. Then, it is possible to integrate subsystems easily into the complete system.

For a particle P moving in time its coordinates are time-dependent too, and they define a trajectory in space. The mathematical representation results in the vector equation $\boldsymbol{x} = \boldsymbol{x}(t)$ equivalent to three scalar equations according to the three degrees of freedom of the particle in the 3D space. The velocity $\boldsymbol{v}(t)$ and the acceleration $\boldsymbol{a}(t)$ of the particle followed by differentiation with respect to time as

$$\boldsymbol{v}(t) = \frac{d^I \boldsymbol{x}(t)}{dt}, \quad \boldsymbol{v}^I(t) = \dot{\boldsymbol{x}}^I(t) = \begin{bmatrix} \dot{x}_1 & \dot{x}_2 & \dot{x}_3 \end{bmatrix}^{\mathrm{T}}, \tag{2.3}$$

$$\boldsymbol{a}(t) = \frac{d^I \boldsymbol{v}(t)}{dt}, \quad \boldsymbol{a}^I(t) = \dot{\boldsymbol{v}}^I(t) = \ddot{\boldsymbol{x}}(t) = \begin{bmatrix} \ddot{x}_1 & \ddot{x}_2 & \ddot{x}_3 \end{bmatrix}^{\mathrm{T}}. \tag{2.4}$$

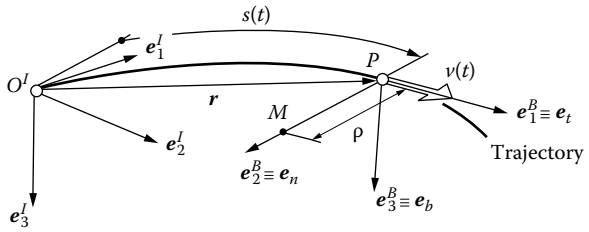


FIGURE 2.3 Trajectory of particle *P*.

The upper right index refers to the frame of reference in which the operations, in particular the differentiation, have to be executed. In the inertial frame *I*, the differentiation of vectors is just performed by differentiation of the scalar coordinates.

The motion of a particle *P* on a curvilinear trajectory in space may be shown in the moving frame *B* too. The position of the point is uniquely identified by the arc length $s(t)$ as a generalized coordinate (Figure 2.3). Then, the position vector $\boldsymbol{r}$ is a function of the arc length, $\boldsymbol{r} = \boldsymbol{r}(s)$. For the velocity and acceleration vector, it yields, see, for example, Magnus and Müller [9],

$$\boldsymbol{v}(t) = \frac{d^I \boldsymbol{r}(t)}{dt} = \frac{d^I \boldsymbol{r}(\mathrm{s})}{ds}\frac{ds}{dt} = v\boldsymbol{e}_t, v = \dot{s},$$

$$\boldsymbol{v}^{\mathrm{B}}(t) = \begin{bmatrix} \dot{s} & 0 & 0 \end{bmatrix}^{\mathrm{T}}, \tag{2.5}$$

$$\boldsymbol{a}(t) = \frac{d^I \boldsymbol{v}(t)}{dt} = a_t \boldsymbol{e}_t + a_n \boldsymbol{e}_n = \dot{v}\boldsymbol{e}_t + \frac{v^2}{\rho}\boldsymbol{e}_n,$$

$$\boldsymbol{a}^{\mathrm{B}}(t) = \begin{bmatrix} \ddot{s} & s^2/\rho & 0 \end{bmatrix}^{\mathrm{T}}, \tag{2.6}$$

where ρ is the curvature of the trajectory in point *P*. Further, it is

$v = \dot{s}$ is the tangent velocity

$a_t = \dot{v} = \ddot{s}$ is the tangent acceleration

$a_n = v^2/\rho = \dot{s}^2/\rho$ is the normal or centripedal acceleration

Special cases of the general motion in space are the motion in a straight line ($\rho \to \infty$), and the motion on a circle (ρ = constant). Often the functions $s(t)$, $\dot{s}(t)$, $\ddot{s}(t)$, $\dot{s}(s)$, $\ddot{s}(s)$, $\ddot{s}(\dot{s})$ are depicted in kinematical diagrams for graphical visualization of the motion along a track.

The translational motion of a rigid body K_i is completely described by the general relations for a particle applied to the center of mass C_i of the rigid body and its position vector $\boldsymbol{r}_i$ (see Figure 2.4).

The rotational motion of a rigid body K_i follows from the relative position of two frames, where one of them is a body-fixed frame. For coinciding origins, an arbitrary position of the body-fixed frame *i* relative to the inertial frame *I* is uniquely defined by three rotation angles according to the three rotational degrees of freedom of a rigid body in space. Both systems are related to each other by three elementary rotations performed successively around different base vectors using three rotation angles. If, for example, the frame *i* is revolved around the coinciding three axes of frame

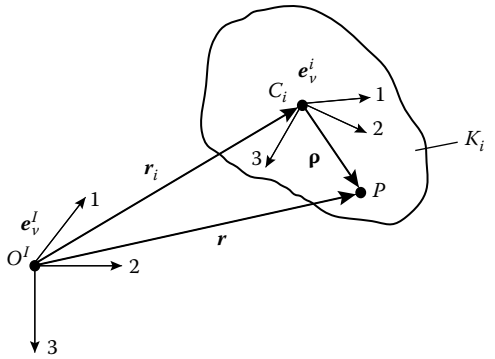


FIGURE 2.4 Position of a rigid body in inertial frame *I*.

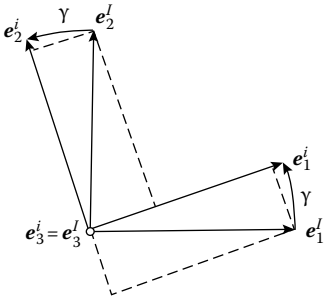


FIGURE 2.5 Elementary rotation from frame *I* to frame *i* around three axes.

I and *i* by the angle γ, the relation of the corresponding base vectors is given by the matrix $\boldsymbol{S}^{Ii}$ as shown in Figure 2.5 and reads as

$$\begin{bmatrix} e_1^I \\ e_2^I \\ e_3^I \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}^{Ii}=\gamma_3} \begin{bmatrix} e_1^i \\ e_2^i \\ e_3^i \end{bmatrix}. \tag{2.7}$$

The row ν of the elementary rotation matrix γ_3 is composed of the coordinates of the base vector $\boldsymbol{e}_\upsilon^I$ in frame *i*. The corresponding matrices for positive rotations around the remaining axes read as

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}, \quad \boldsymbol{\beta}_2 = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}, \tag{2.8}$$

where the elementary rotation matrices are characterized by the name of the rotation angle while the index defines the axis of rotation. There are numerous possibilities to choose the name of the angle and the sequence of the rotation axes used which is not commutative. In vehicle dynamics, the Cardan angles α, β, γ are often used (see Figure 2.6), which are different from the well-known Euler angles ψ, ϑ, φ.

The resulting rotation matrices $\boldsymbol{S}^{Ii}$, which present the relation between frames *I* and *i*, are obtained by the corresponding matrix multiplications:

$$\boldsymbol{S}^{Ii}(\alpha,\beta,\gamma) = \boldsymbol{\alpha}_1 \boldsymbol{\beta}_2 \boldsymbol{\gamma}_3, \quad S^{Ii}(\psi,\vartheta,\varphi) = \boldsymbol{\psi}_3 \boldsymbol{\vartheta}_1 \boldsymbol{\varphi}_3. \tag{2.9}$$

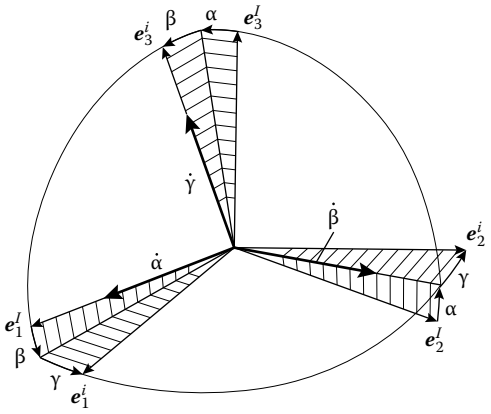


FIGURE 2.6 Spatial rotation from frame *I* to frame *i* by Cardan angles α, β, γ.

Since matrix products are not commutative, the sequence of the elementary rotations has to be strongly observed. The Cardan and Euler angles, respectively, are defined by successive rotations around the 1, 2, 3 axes and 3, 1, 3 axes, respectively, starting from the inertial frame *I*. The sequence of the elementary rotations is uniquely identified by the sequence of the indices of the elementary rotation matrices as shown in 2.9.

The rotation matrices are orthogonal matrices:

$$\left(\boldsymbol{S}^{\boldsymbol{Ii}}\right)^{-1} = \left(\boldsymbol{S}^{\boldsymbol{Ii}}\right)^{\mathrm{T}} = \boldsymbol{S}^{\boldsymbol{iI}}\ , \quad det\boldsymbol{S} = +1, \tag{2.10}$$

where the inversion is also represented by the exchange of the upper indices. The inverse rotation matrix $\boldsymbol{S}^{iI}$ is simply found by transposition of the original rotation matrix $\boldsymbol{S}^{Ii}$.

Using Cardan angles, the rotation matrix reads explicitly as

$$\boldsymbol{S}^{Ii}\left(\alpha,\beta,\gamma\right) = \begin{bmatrix} c\beta c\gamma & -c\beta s\gamma & s\beta \\ c\alpha s\gamma + s\alpha s\beta c\gamma & c\alpha c\gamma - s\alpha s\beta s\gamma & -s\alpha c\beta \\ s\alpha s\gamma - c\alpha s\beta c\gamma & s\alpha c\gamma + c\alpha s\beta s\gamma & c\alpha c\beta \end{bmatrix}, \tag{2.11}$$

where the abbreviations *c* and *s* stand for cos and sin, respectively. In application, often, small rotations are found, $\alpha, \beta, \gamma \ll 1$, resulting in the linearized rotation matrix

$$\boldsymbol{S}^{Ii}\left(\alpha,\beta,\gamma\right) = \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix}. \tag{2.12}$$

This result is also obtained if the elementary rotation matrices are linearized and multiplied with each other. Due to the vector property of small rotations, the sequence of the multiplications has no longer to be considered. For small rotations, the Cardan angles may be assigned directly to the rotational motions around the body-fixed axes (Figure 2.7). In vehicle engineering, the following notations are used:

α, roll motion
β, pitch motion
γ, yaw motion

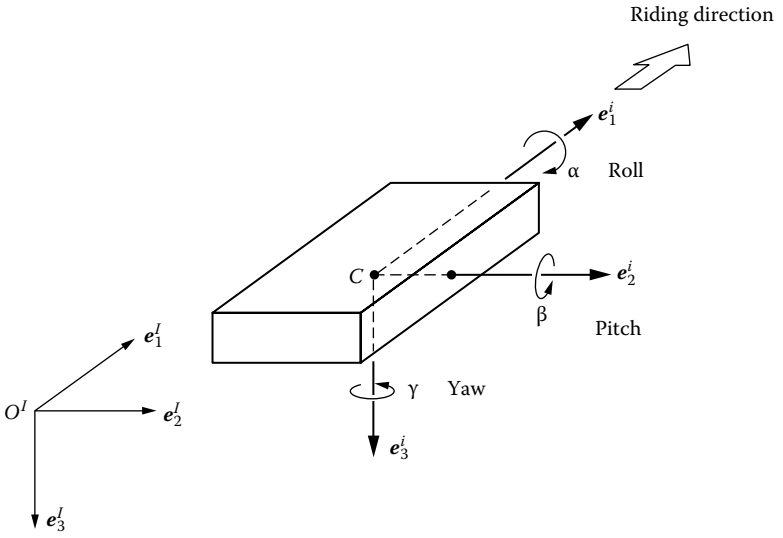


FIGURE 2.7 Notation of rotational vehicle motions.

The coordinates of a vector $\boldsymbol{x}$ read differently for different frames. The relation between the coordinates $\boldsymbol{x}^i$ in frame i and the coordinates $\boldsymbol{x}^I$ in frame I is given by the transformation law for vector coordinates as

$$\boldsymbol{x}^i = \boldsymbol{S}^{iI}\boldsymbol{x}^I \quad \text{and} \quad \boldsymbol{x}^I = \boldsymbol{S}^{Ii}\boldsymbol{x}^i, \tag{2.13}$$

respectively, which is easily proven by 2.10. Please observe that the same indices appear in both forms in a neighboring sequence. This property is often helpful in applications. Further, it has to be pointed out that the rotation matrix $\boldsymbol{S}$ is a function of time, $\boldsymbol{S} = \boldsymbol{S}(t)$, which has to be considered for time derivations.

The position of a rigid body K_i in the inertial frame is uniquely described by the position quantities $\{\boldsymbol{r}_i, \boldsymbol{S}^{Ii}\}$, which characterize the body-fixed frame $\left\{C^i, \boldsymbol{e}_\upsilon^i\right\}$. During motion, the position quantities are functions of time. Thus, the position coordinates of an arbitrary particle P of the rigid body K_i read in the inertial frame I as

$$\boldsymbol{r}^I(t) = \boldsymbol{r}_i^I(t) + \boldsymbol{\rho}^I(t), \quad \boldsymbol{\rho}^I(t) = \boldsymbol{S}^{Ii}(t)\boldsymbol{\rho}^i, \tag{2.14}$$

where it yields in the body-fixed frame ρ^i = constant (see also Figure 2.4).

The motion of a rigid body K_i will be now presented in the inertial frame I, too. The change of the position of its particle P with respect to time relative to frame I is found by differentiation of 2.14 as

$$\dot{\boldsymbol{r}}^I(t) = \dot{\boldsymbol{r}}_i^I(t) + \dot{\boldsymbol{S}}^{Ii}(t)\boldsymbol{\rho}^i = \dot{\boldsymbol{r}}_i^I(\mathrm{t}) + \dot{\boldsymbol{S}}^{Ii}(t)\boldsymbol{S}^{\mathrm{iI}}(t)\boldsymbol{\rho}^I(t). \tag{2.15}$$

The first term on the right-hand side represents the translational velocity of the origin C_i of the body-fixed frame I. The second term is obviously related to the rotation of the body-fixed frame and represents the body's rotation. This term will now be discussed in more detail. The matrix product

$\dot{\boldsymbol{S}}(t)\boldsymbol{S}^{\mathrm{T}}(t)$ is screw symmetric, that is, $[\bullet] = -[\bullet]^{\mathrm{T}}$, which follows immediately from the differentiation of the orthogonality condition $\boldsymbol{S}(t)\boldsymbol{S}^{\mathrm{T}}(t) = \boldsymbol{E}$ according to 2.10:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left[\boldsymbol{S}(t)\boldsymbol{S}^{\mathrm{T}}(t)\right] &= \dot{\boldsymbol{S}}(t)\boldsymbol{S}^{\mathrm{T}}(t) + \boldsymbol{S}(t)\dot{\boldsymbol{S}}^{\mathrm{T}}(t) \\ &= \dot{\boldsymbol{S}}(t)\boldsymbol{S}^{\mathrm{T}}(t) + \left[\dot{\boldsymbol{S}}(t)\boldsymbol{S}^{\mathrm{T}}(t)\right]^{\mathrm{T}} = 0. \end{aligned} \tag{2.16}$$

The matrix product $[\bullet]$ will be abbreviated by the symbol $\tilde{\boldsymbol{\omega}}(t)$ and identified by the corresponding three coordinates $\boldsymbol{\omega}_{\mathrm{v}} = \boldsymbol{\omega}_{\mathrm{v}}(t)$ as follows:

$$\dot{\boldsymbol{S}}^{Ii}\left(\boldsymbol{S}^{Ii}\right)^{\mathrm{T}} = \dot{\boldsymbol{S}}^{Ii}\boldsymbol{S}^{iI} = \tilde{\boldsymbol{\omega}}_{Ii}^{I}(t) := \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}, \quad \boldsymbol{\omega}_{Ii}^{I} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix}. \tag{2.17}$$

Both quantities, the screw-symmetric tensor $\tilde{\boldsymbol{\omega}}_{Ii}^{I}$ and the corresponding rotational velocity vector $\boldsymbol{\omega}_{Ii}^{I}$, respectively, describe the rotational motion of system i or body K_i, respectively, relative to the inertial frame I. The upper indices indicate that both quantities are represented in the inertial frame I. If there is no chance for mixing up the frames, the upper and lower index I is simply deleted. The screw-symmetric tensor corresponding with a vector $(\bullet)$ is identified by the symbol $(\tilde{\bullet})$ and it replaces the vector product

$$\tilde{\boldsymbol{\omega}}\boldsymbol{\rho} \equiv \boldsymbol{\omega} \times \boldsymbol{\rho}. \tag{2.18}$$

In coordinates, in any frame, one gets accordingly

$$\tilde{\boldsymbol{\omega}}\boldsymbol{\rho} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} \omega_2\rho_3 - \omega_3\rho_2 \\ \omega_3\rho_1 - \omega_1\rho_3 \\ \omega_1\rho_2 - \omega_2\rho_1 \end{bmatrix}. \tag{2.19}$$

This notation of the vector product is most valuable for numerical computations since the vector product is not defined in matrix calculus.

The rotational velocity vector $\boldsymbol{\omega}_{Ii}^{i}$ in the body-fixed frame i follows from transformation or direct evaluation, respectively. The application of transformation 2.13 to 2.19 results in

$$\tilde{\boldsymbol{\omega}}_{Ii}^{i} = \boldsymbol{S}^{iI}\tilde{\boldsymbol{\omega}}_{Ii}^{I}\boldsymbol{S}^{Ii} = \boldsymbol{S}^{iI}\left(\dot{\boldsymbol{S}}^{Ii}\boldsymbol{S}^{iI}\right)\boldsymbol{S}^{Ii} = \boldsymbol{S}^{iI}\dot{\boldsymbol{S}}^{Ii} = \left(\boldsymbol{S}^{Ii}\right)^{T}\dot{\boldsymbol{S}}^{Ii}. \tag{2.20}$$

The first and second terms of 2.20 represent the transformation law for tensor coordinates where the same indices appear again in a neighboring sequence, the first and last terms show the direct evaluation. The corresponding vector to 2.20 is $\boldsymbol{\omega}_{Ii}^{i} = \boldsymbol{S}^{iI}\boldsymbol{\omega}_{Ii}^{I}$, where the transformation law for vector coordinates has to be used again.

From 2.15 and 2.17, it follows for rigid body K_i

$$\boldsymbol{v}^{I}(t) = \boldsymbol{v}_i^{I}(t) + \tilde{\boldsymbol{\omega}}_{Ii}^{I}(t)\boldsymbol{\rho}^{I}(t). \tag{2.21}$$

Considering 2.18, this is the relation for rigid body kinematics well known from each mechanics textbook as

$$\boldsymbol{v}(t) = \boldsymbol{v}_i(t) + \boldsymbol{\omega}_{Ii}(t) \times \rho(t). \tag{2.22}$$

The relations 2.21 or 2.22, respectively, represent the motion of a rigid body composed by an absolute translational velocity $\boldsymbol{v}_i$ of the body-fixed reference point $O^i = C^i$, and a rotation with the angular velocity $\boldsymbol{\omega}_{Ii}$. The fundamental kinematical quantities $\{\boldsymbol{v}_i, \boldsymbol{\omega}_{Ii}\}$ are also denoted as twist characterizing uniquely the motion of a rigid body.

2.2.3 Kinematics of a Rigid Body in a Moving Reference Frame

From a mathematical point of view, the kinematical description of the motion of a rigid body is most convenient in the inertial frame I resulting in a more simple representation of the fundamental laws of mechanics. In engineering applications, however, a moving reference frame R related to the vehicle or the guideway, respectively, turns out to be more adequate. The frame R allows a problem-oriented choice of the coordinates and an efficient description of the forces and torques acting on the system. Moving reference frames are also useful in experiments since many measurement data are not related to the inertial frame. The choice of the reference frame R depends on the problem under consideration. In many cases, the frame R characterizes the large nonlinear reference motion of a vehicle while the small deviation from the reference motion results in linear kinematical relations even for rotations.

In the following, the motion of a rigid body is represented in a moving reference frame R the motion of which is known in the inertial frame I by the position vector $\boldsymbol{r}_R(t)$ of its origin, and rotation matrix $\boldsymbol{S}^{IR}(t)$ (see also Figure 2.8). This means that the translational and angular guidance velocities are also known, $\boldsymbol{v}_R^I = \dot{\boldsymbol{r}}_R^I$, $\tilde{\omega}_{IR}^I = \dot{\boldsymbol{S}}^{IR}\boldsymbol{S}^{RI}$ according to 2.3 and 2.17. Considering Figure 2.8, the absolute position quantities $\{\boldsymbol{r}_i, \boldsymbol{S}^{Ii}\}$ of the rigid body read as

$$\boldsymbol{r}_i^I(t) = \boldsymbol{r}_R^I(t) + \boldsymbol{S}^{IR}(t)\boldsymbol{r}_{Ri}^R(t), \tag{2.23}$$

$$\boldsymbol{S}^{Ii}(t) = \boldsymbol{S}^{IR}(t)\boldsymbol{S}^{Ri}(t). \tag{2.24}$$

The absolute motion $\{\boldsymbol{v}_i, \omega_{Ii}\}$ of the rigid body K_i is now found by formal differentiation in frame I and subsequent transformation in frame R as

$$\boldsymbol{v}_i^R(t) = \boldsymbol{v}_R^R(t) + \tilde{\boldsymbol{\omega}}_{IR}^R(t)\boldsymbol{r}_{Ri}^R(t) + \dot{\boldsymbol{r}}_{Ri}^R(t), \tag{2.25}$$

$$\boldsymbol{\omega}_{Ii}^R = \boldsymbol{\omega}_{IR}^R + \boldsymbol{\omega}_{Ri}^R. \tag{2.26}$$

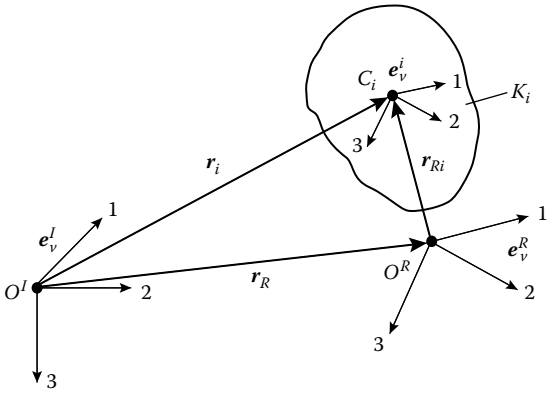


FIGURE 2.8 Position of a rigid body in reference frame R.

Due to the rotation of the frame R, the guidance motion and the relative motion characterized by the indices IR and Ri, respectively, are not simply added but there appears an additional term in 2.25. This implies the well-known law of differentiation in a rotating frame

$$\frac{\mathrm{d}^I}{\mathrm{d}t}\boldsymbol{r}(t) = \frac{\mathrm{d}^R}{\mathrm{d}t}\boldsymbol{r}(t) + \boldsymbol{\omega}_{IR}(t) \times \boldsymbol{r}(t). \tag{2.27}$$

By formal differentiation with respect to the inertial frame indicated by [*] or application of 2.27 to 2.25 and 2.26, one gets finally the absolute translational and rotational acceleration of the rigid body K_i again written in the reference frame R as

$$\boldsymbol{a}_i^R(t) = \boldsymbol{v}^{*R}_R + \left(\dot{\tilde{\boldsymbol{\omega}}}^R_{IR} + \tilde{\boldsymbol{\omega}}^R_{IR}\,\tilde{\boldsymbol{\omega}}^R_{IR}\right)\boldsymbol{r}^R_{Ri} + 2\tilde{\boldsymbol{\omega}}^R_{IR}\,\dot{\boldsymbol{r}}^R_{Ri} + \ddot{\boldsymbol{r}}^R_{Ri}, \tag{2.28}$$

$$\boldsymbol{\alpha}^R_{Ii}(t) = \dot{\boldsymbol{\omega}}^R_{IR} + \tilde{\boldsymbol{\omega}}^R_{IR}\,\boldsymbol{\omega}^R_{Ri} + \dot{\boldsymbol{\omega}}^R_{Ri}. \tag{2.29}$$

Thus, in addition to the guidance and relative acceleration the Coriolis acceleration with the characteristic factor 2 is found for translations.

2.2.4 Kinematics of Multi-Body Systems

So far only one free rigid body K_i was considered the position of which is uniquely described in the inertial frame I as

$$\boldsymbol{r}_i^I = \left[r_{i1}\, r_{i2}\, r_{i3}\right]^{\mathrm{T}}, \quad \boldsymbol{S}^{Ii} \equiv \boldsymbol{S}_i = \boldsymbol{S}_i(\alpha_i, \beta_i, \gamma_i). \tag{2.30}$$

There are six position coordinates, which are summarized in a 6×1 column matrix, simply called local position vector, as

$$\boldsymbol{x}_i(t) = \left[r_{i1}\, r_{i2}\, r_{i3}\;\; \alpha_i\, \beta_i\, \gamma_i\right]^{\mathrm{T}}. \tag{2.31}$$

For a free multi-body system consisting of p dismantled rigid bodies K_i, $i = 1(1)p$, there remain $6p$ position coordinates resulting in a $6p \times 1$ global position vector of an unconstrained system

$$\boldsymbol{x}(t) = [\boldsymbol{x}_1^{\mathrm{T}} \ldots \boldsymbol{x}_p^{\mathrm{T}}]^{\mathrm{T}}. \tag{2.32}$$

Assembling the free system, there appear constraints between the position coordinates and their derivatives. In realistic models of vehicles, only holonomic constraints are found restricting the motion of the position coordinates by geometric or integrable kinematic constraints. These constraints are implicitly described by algebraical equations, which may be time-dependent (rheonomic), too:

$$\varphi_j(\boldsymbol{x}, t) = 0, \quad j = 1(1)q. \tag{2.33}$$

Due to q constraints, there remain f linear-independent position coordinates characterizing $f = 6p - q$ degrees of freedom. The f-independent position coordinates are also called generalized coordinates and may be summarized in a $f \times 1$ column matrix as global position vector of the constraint system:

$$\boldsymbol{y}(t) = \left[\, y_1 \ldots y_f \,\right]^{\mathrm{T}}. \tag{2.34}$$

By 2.33 and 2.34, the vector $\boldsymbol{x}$ is an explicit function of the f-generalized coordinates representing the constraints explicitly,

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{y}, t). \tag{2.35}$$

The choice of generalized coordinates is not unique. For example, some of the local position coordinates (absolute coordinates) or differences between local coordinates (relative coordinates) may be chosen as generalized coordinates. However, there exists a unique relation between different sets of generalized coordinates represented by a regular, time-invariant $f \times f$ matrix $\boldsymbol{T}$ resulting in the transformation

$$\boldsymbol{y}(t) = \boldsymbol{T}\,\overline{\boldsymbol{y}}(t), \tag{2.36}$$

where $\boldsymbol{y}$ and $\overline{\boldsymbol{y}}$ are the corresponding global position vectors.

The position variables 2.30 may be rewritten for the whole system as

$$\boldsymbol{r}_i(t) = \boldsymbol{r}_i(y, t), \quad \boldsymbol{S}^{Ii}(t) \equiv \boldsymbol{S}_i(t) = \boldsymbol{S}_i(\boldsymbol{y}, t), \quad i = 1(1)p. \tag{2.37}$$

The corresponding velocity variables $\{\boldsymbol{v}_i, \boldsymbol{\omega}_i\}$, $\boldsymbol{\omega}_i = \boldsymbol{\omega}_{Ii}$ are obtained by differentiation as

$$\boldsymbol{v}_i(t) = \dot{\boldsymbol{r}}_i(t) = \frac{\partial \boldsymbol{r}_i}{\partial \boldsymbol{y}^{\mathrm{T}}}\,\dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{r}_i}{\partial t} = \boldsymbol{J}_{Ti}(\boldsymbol{y}, t)\dot{\boldsymbol{y}} + \overline{\boldsymbol{v}}_i(\boldsymbol{y}, t), \tag{2.38}$$

$$\boldsymbol{\omega}_i(t) = \dot{\boldsymbol{s}}_i(t) = \frac{\partial \boldsymbol{s}_i}{\partial \boldsymbol{y}^{\mathrm{T}}}\,\dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{s}_i}{\partial t} = \boldsymbol{J}_{Ri}(\boldsymbol{y}, t)\dot{\boldsymbol{y}} + \overline{\boldsymbol{\omega}}_i(y, t), \tag{2.39}$$

where $\partial \boldsymbol{s}_i$ describes the 3×1 vector of the infinitesimal rotation following from the rotation matrix analogously to the rotational velocity 2.17 as

$$\partial \tilde{\boldsymbol{s}}_i = \partial \boldsymbol{S}_i \boldsymbol{S}_i^{\mathrm{T}} = \begin{bmatrix} 0 & -\partial s_{i3} & \partial s_{i2} \\ \partial s_{i3} & 0 & -\partial s_{i1} \\ -\partial s_{i2} & \partial s_{i1} & 0 \end{bmatrix}, \quad \partial \boldsymbol{s}_i = \begin{bmatrix} \partial s_{i1} \\ \partial s_{i2} \\ \partial s_{i3} \end{bmatrix}. \tag{2.40}$$

The $3 \times f$ functional or Jacobian matrices $\boldsymbol{J}_{Ti}, \boldsymbol{J}_{Ri}$ of translation and rotation, respectively, identify the relation between the local and the generalized or global coordinates. The formation of these matrices is defined using the rules of matrix multiplication as shown for the translation matrix:

$$\frac{\partial \boldsymbol{r}_i}{\partial \boldsymbol{y}^T} = \partial \boldsymbol{r}_i \left(\frac{1}{\partial \boldsymbol{y}^T} \right) = \boldsymbol{J}_{Ti} = \begin{matrix} \\ \\ \\ \end{matrix} \begin{matrix} \dfrac{\partial r_{i1}}{\partial y_1} & \dfrac{\partial r_{i1}}{\partial y_2} & \cdots & \dfrac{\partial r_{i1}}{\partial y_f} \\ \dfrac{\partial r_{i2}}{\partial y_1} & \dfrac{\partial r_{i2}}{\partial y_2} & \cdots & \dfrac{\partial r_{i2}}{\partial y_f} \\ \dfrac{\partial r_{i3}}{\partial y_1} & \dfrac{\partial r_{i3}}{\partial y_2} & \cdots & \dfrac{\partial r_{i3}}{\partial y_f} \end{matrix}. \tag{2.41}$$

From 2.38 and 2.39, one obtains by a second differentiation the acceleration variables $\{\boldsymbol{a}_i, \alpha_i\}$ depending on the position vector $\boldsymbol{y}$ and its derivatives:

$$\boldsymbol{a}_i(t) = \dot{\boldsymbol{v}}_i(t) = \boldsymbol{J}_{Ti}(\boldsymbol{y}, t)\ddot{\boldsymbol{y}} + \frac{\partial \boldsymbol{v}_i}{\partial \boldsymbol{y}^{\mathrm{T}}}\dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{v}_i}{\partial t}, \tag{2.42}$$

$$\boldsymbol{\alpha}_i(t) = \dot{\boldsymbol{\omega}}_i(t) = \boldsymbol{J}_{Ri}(\boldsymbol{y}, t)\ddot{\boldsymbol{y}} + \frac{\partial \boldsymbol{\omega}_i}{\partial \boldsymbol{y}^{\mathrm{T}}}\dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{\omega}_i}{\partial t}. \tag{2.43}$$

For scleronomic, time-invariant constraints, the partial time derivatives in 2.38, 2.39, 2.42, and 2.43 are vanishing.

In addition to the real motions, the virtual motions are required in the next chapter dealing with dynamics. A virtual motion is defined as an arbitrary, infinitesimally small variation of the position completely compatible with the constraints at any time. Rheonomic constraints are considered to be frozen at the time under consideration. The symbol δ of the virtual motion has the properties

$$\delta\boldsymbol{r} \neq \boldsymbol{0}, \delta t \equiv 0. \tag{2.44}$$

The symbol δ follows the rules of calculus, that is, it yields

$$\delta(c\boldsymbol{r}) = c\delta\boldsymbol{r}, \quad \delta(\boldsymbol{r}_1 + \boldsymbol{r}_2) = \delta\boldsymbol{r}_1 + \delta\boldsymbol{r}_2, \quad \delta\boldsymbol{r}(\boldsymbol{y}) = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{y}^T}\delta\boldsymbol{y}. \tag{2.45}$$

Thus, the virtual motion of a multi-body system reads as

$$\delta\boldsymbol{r}_i = \boldsymbol{J}_{Ti}\delta\boldsymbol{y}, \quad \delta\boldsymbol{s}_i = \boldsymbol{J}_{Ri}\,\delta\boldsymbol{y}, \quad i = 1(1)p. \tag{2.46}$$

This completes the kinematics for rigid body vehicle systems.

2.3 DYNAMICS

For the generation of the equations of motion of multi-body systems, in addition to kinematics, the inertia of the bodies and the acting forces have to be considered. The Newton–Euler approach, also called the synthetic method, uses the free-body diagram resulting in full set of local equations, which may be reduced by the principles of d'Alembert and Jourdain to the equations of motion. The Lagrangian approach, representing the analytical method, is based on energy considerations and the equations of motions are found directly but without any information on the reaction forces.

2.3.1 Inertia Properties

The inertia of a rigid body K_i is characterized by its mass m_i and its inertia tensor $\boldsymbol{I}_{Ci}$. The coordinates of the inertia tensor read in the body-fixed frame $\left\{C_i, \boldsymbol{e}_{\upsilon}^{i}\right\}$ (see Figure 2.4) as

$$\boldsymbol{I}_{Ci}^{i} = \int_{m_i}(\rho^{\mathrm{T}}\rho\boldsymbol{E} - \rho\rho^{\mathrm{T}})\,dm = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}_{C_i} = \text{constant}. \tag{2.47}$$

The vector $\boldsymbol{\rho} \equiv \boldsymbol{\rho}^i = [\rho_1\ \rho_2\ \rho_3]^{\mathrm{T}}$ describes a material point with mass dm with respect to the center of mass C_i and $\boldsymbol{E}$ means the 3×3 identity matrix. The inertia tensor $\boldsymbol{I}_{Ci}^i$ is symmetric and positive definite, and constant in the body-fixed frame.

The coordinates of the inertia tensor depend on the mass distribution and on the choice of the reference frame. For a parallel displacement of the body-fixed frame from the center of mass C_i to an arbitrary body-fixed point O_i characterized by the vector $\boldsymbol{s}$, one gets

$$\boldsymbol{I}_{Oi}^i = \boldsymbol{I}_{Ci}^i + (\boldsymbol{s}^{\mathrm{T}}\boldsymbol{s}\boldsymbol{E} - \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}})m_i. \tag{2.48}$$

Thus, the diagonal elements of an inertia tensor are minimal for the center of mass.

For a homogeneous, purely rotational displacement by the rotation matrix $\boldsymbol{S}^{ii'}$ from frame $\boldsymbol{e}_\upsilon^i$ to $\boldsymbol{e}_\upsilon^{i'}$ around the center of mass the transformation law for tensors applies as

$$\boldsymbol{I}_{Ci}^i = \boldsymbol{S}^{ii'}\boldsymbol{I}_{Ci}^{i'}\boldsymbol{S}^{i'i} \quad \text{or} \quad \boldsymbol{I}_{Ci}^{i'} = \boldsymbol{S}^{i'i}\boldsymbol{I}_{Ci}^i\boldsymbol{S}^{ii'}. \tag{2.49}$$

Please note that the inertia tensor may be time-variant if the frame $\left\{C_i, \boldsymbol{e}_\upsilon^{i'}\right\}$ is not body fixed. This is especially true if the inertial frame is chosen, $i' \equiv I$, due to $\boldsymbol{S}^{iI} = \boldsymbol{S}^{iI}(t)$.

For all reference points, there exists a special body-fixed frame in which the off-diagonal elements of the inertia tensor are vanishing, for example,

$$\boldsymbol{I}_{Ci}^{i'} = \mathrm{diag}\left[I_1\ \ I_2\ \ I_3\right] = \text{constant}. \tag{2.50}$$

The remaining diagonal elements I_υ are called principal moments of inertia with reference to C_i and the corresponding axes are the principal inertia axes. Both quantities follow from the eigenvalue problem:

$$(I_\upsilon \boldsymbol{E} - \boldsymbol{I}_{Ci}^{i'})\boldsymbol{x}_\upsilon = \boldsymbol{0}. \tag{2.51}$$

Thus, the principal moments of inertia are the eigenvalues of the matrix $\boldsymbol{I}_{Ci}^{i'}$ and the eigenvectors $\boldsymbol{x}_\upsilon = \boldsymbol{e}_\upsilon^i$ define the principal inertia axes which have to be unit vectors $\boldsymbol{x}_\upsilon^{\mathrm{T}}\boldsymbol{x}_\upsilon = 1$.

2.3.2 Newton–Euler Equations

The synthetic method is based on the laws of Newton (1687) and Euler (1758) relating the translational motion represented by the momentum $\boldsymbol{p}$ of a body K to the sum of the external forces $\boldsymbol{f}$ and the rotational motion represented by the moment of momentum $\boldsymbol{h}_O$ to the sum of the external torques $\boldsymbol{l}_O$,

$$\frac{\mathrm{d}^I}{\mathrm{d}t}\boldsymbol{p} = \boldsymbol{f}, \quad \frac{\mathrm{d}^I}{\mathrm{d}t}\boldsymbol{h}_O = \boldsymbol{l}_O. \tag{2.52}$$

The time derivatives of the momentum $\boldsymbol{p}$ and the moment of momentum $\boldsymbol{h}_O$ have to be evaluated in the inertial frame I. The common reference point O of the moment of momentum and the resulting external torque may be an inertially fixed point like the origin of the inertial frame, $O \equiv O_I$, or the moving center of mass of the body, $O \equiv C$.

The fundamental laws (2.52) will now be applied to the rigid body K_i, $i = 1(1)p$, of a multi-body system and appropriate frames are chosen. First of all, the bodies K_i are dismantled and the constraints are replaced by reaction forces acting then externally on the bodies involved in the same

size but with opposite sign according to the counteraction principle (action = reaction). Further, the center of mass is used as reference point for all bodies, $O \equiv C_i$.

In the inertial frame, I momentum and moment of momentum for a rigid body K_i using the inertia properties m_i, $\boldsymbol{I}_{Ci}$ read as

$$\boldsymbol{p}_i^I = m_i \boldsymbol{v}_{Ci}^I, \quad m_i = \text{constant}, \tag{2.53}$$

$$\boldsymbol{h}_{Ci}^I = \boldsymbol{I}_{Ci}^I \boldsymbol{\omega}_i^I, \quad \boldsymbol{I}_{Ci}^I = \boldsymbol{I}_{Ci}^I(t), \tag{2.54}$$

where $\boldsymbol{v}_{Ci}^I$ and $\boldsymbol{\omega}_i^I$ mean absolute velocities. Introducing 2.53 and 2.54 in 2.52 and omitting the index C, one finally gets Newton's and Euler's equations:

$$m_i \dot{\boldsymbol{v}}_i^I = \boldsymbol{f}_i^I, \quad m_i = \text{constant}, \tag{2.55}$$

$$\boldsymbol{I}_i^I \dot{\boldsymbol{\omega}}_i^I + \tilde{\boldsymbol{\omega}}_i^I \boldsymbol{I}_i^I \boldsymbol{\omega}_i^I = \boldsymbol{l}_i^I, \quad \boldsymbol{I}_i^I = \boldsymbol{I}_i^I(t). \tag{2.56}$$

In a second step, these equations are transformed in a body-fixed frame resulting in

$$m_i \dot{\boldsymbol{v}}_i^i = \boldsymbol{f}_i^i, \; m_i = \text{constant}, \tag{2.57}$$

$$\boldsymbol{I}_i^i \dot{\boldsymbol{\omega}}_i^i + \tilde{\boldsymbol{\omega}}_i^i \boldsymbol{I}_i^i \boldsymbol{\omega}_i^i = \boldsymbol{l}_i^i, \boldsymbol{I}_i^i = \text{constant}. \tag{2.58}$$

Equations 2.55 and 2.56, and 2.57 and 2.58 look completely identical. If there is only one body like in gyrodynamics, then 2.58 is preferable due to the time invariance of the inertia tensor. In multibody dynamics, however, this advantage is fading. Equation 2.58 is also known as Euler's equation of gyrodynamics. It can be found with the moment of momentum given in the body-fixed frame from 2.52 directly using the law of differentiation in a rotating frame (2.27),

$$\frac{\mathrm{d}^I}{\mathrm{d}t} \boldsymbol{h}_i^i = \dot{\boldsymbol{h}}_i^i + \tilde{\boldsymbol{\omega}}_i^i \boldsymbol{h}_i^i = \boldsymbol{l}_i^i, \; \boldsymbol{h}_i^i = \boldsymbol{I}_i^i \boldsymbol{\omega}_i^i. \tag{2.59}$$

Finally, in an arbitrarily moving reference frame R Newton's and Euler's equations are also available (see, e.g., Schiehlen and Eberhard [10]):

$$m_i \ddot{\boldsymbol{r}}_{Ri} + m_i \left[\overset{**}{\boldsymbol{r}}_R + (\dot{\tilde{\boldsymbol{\omega}}}_R + \tilde{\boldsymbol{\omega}}_R \, \tilde{\boldsymbol{\omega}}_R) \boldsymbol{r}_{Ri} + 2 \tilde{\boldsymbol{\omega}}_R \, \dot{\boldsymbol{r}}_{Ri} \right] = \boldsymbol{f}_i, \tag{2.60}$$

$$\boldsymbol{I}_i \dot{\boldsymbol{\omega}}_{Ri} + \tilde{\boldsymbol{\omega}}_{Ri} \, \boldsymbol{I}_i \boldsymbol{\omega}_{Ri} + \left[\boldsymbol{I}_i \boldsymbol{\omega}_R + \tilde{\boldsymbol{\omega}}_R \, \boldsymbol{I}_i \, \boldsymbol{\omega}_R + \tilde{\boldsymbol{\omega}}_R \, \boldsymbol{\omega}_{Ri} \, sp \, \boldsymbol{I}_i + 2 \tilde{\boldsymbol{\omega}}_{Ri} \, I_i \, \boldsymbol{\omega}_R \right] = \boldsymbol{l}_i. \tag{2.61}$$

Now, the coordinates of all vectors and tensors are related to the reference frame R where $[\overset{**}{\boldsymbol{r}}_R]$ means that the second-time derivation has to be made before in the inertial frame. As a matter of fact, a large number of additional inertia forces and torques appear due to the relative motion.

The Newton–Euler equations represent a set of $6p$ scalar equations for $6p$ unknowns, which are composed of unknown velocity and position variables and unknown reaction forces and torques. In an unconstrained system, reactions do not exit, that is, there are $6p$ ordinary differential equations (ODEs) to be solved. In a completely constrained system, motion does not occur at all, that is, altogether $6p$ algebraical equations have to be solved. In vehicle dynamics, due to a certain number of constraints between the bodies, motions and reactions appear featuring a set of differential-algebraical equations (DAEs). However, by the principles of dynamics, a minimal set of f ODEs can be found facilitating the solution and simulation of the problem.

2.3.3 Principles of d'Alembert and Jourdain

Equations of motion represent a minimal set of ODEs. They can be found from the Newton–Euler equations by elimination of the reaction forces and torques. This is achieved computationally efficient by the principles of dynamics considering the virtual work of a constrained multi-body system. For this purpose, the external forces acting on the dismantled bodies of the system are subdivided into applied forces $\boldsymbol{f}_i^{(a)}$ and torques $\boldsymbol{l}_i^{(a)}$ as well as reaction forces $\boldsymbol{f}_i^{(r)}$ and torques $\boldsymbol{l}_i^{(r)}$. The latter ones do not contribute to the virtual work of the system

$$\delta W^r = \sum_{i=1}^{p} (\boldsymbol{f}_i^{(r)\mathrm{T}}\, \delta \boldsymbol{r}_i + \boldsymbol{l}_i^{(r)\mathrm{T}}\, \delta \boldsymbol{s}_i) = 0, \tag{2.62}$$

where the virtual motions $\delta \boldsymbol{r}_i$, $\delta \boldsymbol{s}_i$ are known from 2.46. Equation 2.62 can be interpreted as a generalized orthogonality condition. For this purpose, in addition to the generalized coordinates $\boldsymbol{y}$, generalized reaction forces g_j, $j = 1(1)q$ are introduced and summarized in a $q \times 1$ vector as

$$\boldsymbol{g}(t) = \left[g_1 \dots g_q \right]^{\mathrm{T}}. \tag{2.63}$$

The number of generalized constraint forces is determined by the number q of constraints. The local constraint forces and torques follow from the implicit constraint equations (2.33) as

$$\boldsymbol{f}_i^{(r)\mathrm{T}} = \sum_{j=1}^{q} g_j \frac{\partial \varphi_j}{\partial \boldsymbol{r}_i^{\mathrm{T}}} = \sum_{j=1}^{q} g_j \frac{\partial \varphi_j}{\partial \boldsymbol{x}^{\mathrm{T}}} \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{r}_i^{\mathrm{T}}}, \tag{2.64}$$

$$\boldsymbol{l}_i^{(r)\mathrm{T}} = \sum_{j=1}^{q} g_j \frac{\partial \varphi_j}{\partial \boldsymbol{s}_i^{\mathrm{T}}} = \sum_{j=1}^{q} g_j \frac{\partial \varphi_j}{\partial \boldsymbol{x}^{\mathrm{T}}} \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{s}_i^{\mathrm{T}}}, \quad i = 1(1)p. \tag{2.65}$$

In matrix notation, the $3 \times q$ Jacobians $\boldsymbol{F}_{Ti}$, $\boldsymbol{F}_{Ri}$ are found from 2.64 and 2.65, and the condition 2.62 is rewritten as

$$\begin{aligned} \delta W^r &= \boldsymbol{g}^{\mathrm{T}} \sum_{i=1}^{p} (\boldsymbol{F}_{Ti}^{\mathrm{T}}\, \delta \boldsymbol{r}_i + \boldsymbol{F}_{Ri}^{\mathrm{T}} \delta \boldsymbol{s}_i) \\ &= \boldsymbol{g}^{\mathrm{T}} \sum_{i=1}^{p} (\boldsymbol{F}_{Ti}^{\mathrm{T}}\, \boldsymbol{J}_{Ti} + \boldsymbol{F}_{Ri}^{\mathrm{T}}\, \boldsymbol{J}_{Ri})\, \delta \boldsymbol{y} = 0. \end{aligned} \tag{2.66}$$

Finally, the global $6p \times q$ distribution matrix $\bar{\boldsymbol{Q}}$, and the global $6p \times f$ Jacobian matrix $\bar{\boldsymbol{J}}$ are introduced:

$$\bar{\boldsymbol{Q}} = [\boldsymbol{F}_{T1}^{\mathrm{T}} \ldots \boldsymbol{F}_{Tp}^{\mathrm{T}}\, \boldsymbol{F}_{R1}^{\mathrm{T}} \ldots \boldsymbol{F}_{Rp}^{\mathrm{T}}]^{\mathrm{T}},$$

$$\bar{\boldsymbol{J}} = [\boldsymbol{J}_{T1}^{\mathrm{T}} \ldots \boldsymbol{J}_{Tp}^{\mathrm{T}}\, \boldsymbol{J}_{R1}^{\mathrm{T}} \ldots \boldsymbol{J}_{Rp}^{\mathrm{T}}]^{\mathrm{T}}. \tag{2.67}$$

Then, one gets from 2.66 simply

$$\bar{\boldsymbol{Q}}^{\mathrm{T}} \bar{\boldsymbol{J}} = \bar{\boldsymbol{J}}^{\mathrm{T}} \bar{\boldsymbol{Q}} = \boldsymbol{0}, \tag{2.68}$$

which clearly shows the generalized orthogonality between motion and constraint. The orthogonality condition or the vanishing virtual work, respectively, is independent from the coordinates chosen, and it is valid for all constrained mechanical systems.

d'Alembert's principle (1743) follows now from the Newton–Euler equations (2.55) and (2.56) after subdividing the external forces

$$\boldsymbol{f}_i = \boldsymbol{f}_i^{(e)} + \boldsymbol{f}_i^{(r)}, \quad \boldsymbol{l}_i = \boldsymbol{l}_i^{(e)} + \boldsymbol{l}_i^{(r)}, \tag{2.69}$$

and considering the orthogonality (2.62) as

$$\sum_{i=1}^{p} [(m_i \dot{\boldsymbol{v}}_i - \boldsymbol{f}_i^{(e)})^{\mathrm{T}} \delta \boldsymbol{r}_i + (\boldsymbol{I}_i \dot{\boldsymbol{\omega}}_i + \tilde{\boldsymbol{\omega}}_i \boldsymbol{I}_i\, \boldsymbol{\omega}_i - \boldsymbol{l}_i^{(e)})^{\mathrm{T}} \delta \boldsymbol{s}_i] = 0. \tag{2.70}$$

Obviously, the reaction forces are eliminated in 2.70.

Analogously, Jourdain's principle (1908) can be stated which is based on the fact that the virtual power of the reaction forces is vanishing, too,

$$\delta P^r = \sum_{i=1}^{p} \left[\boldsymbol{f}_i^{(r)\mathrm{T}}\, \delta' \boldsymbol{v}_i + \boldsymbol{l}_i^{(r)\mathrm{T}}\, \delta' \boldsymbol{\omega}_i \right] = 0. \tag{2.71}$$

The virtual velocities $\delta' \boldsymbol{v}_i$, $\delta' \boldsymbol{\omega}_i$ are arbitrary, infinitesimal small variations of the velocities completely compatible with the constraints at any time and at any position. Thus, it yields

$$\delta' \boldsymbol{v}_i \neq 0\,, \quad \delta' \boldsymbol{\omega}_i \neq 0\,, \quad \delta' \boldsymbol{r}_i \equiv 0\,, \quad \delta' \boldsymbol{s}_i \equiv 0\,, \quad \delta' t \equiv 0. \tag{2.72}$$

Moreover, the symbol δ' follows the rules of calculus. Then, it remains as Jourdain's principle as

$$\sum_{i=1}^{p} [(m_i \dot{\boldsymbol{v}}_i - \boldsymbol{f}_i^{(e)})^{\mathrm{T}} \delta' \boldsymbol{v}_i + (\boldsymbol{I}_i \dot{\boldsymbol{\omega}}_i + \tilde{\boldsymbol{\omega}}_i \boldsymbol{I}_i \boldsymbol{\omega}_i - \boldsymbol{l}_i^{(e)})^{\mathrm{T}} \delta' \boldsymbol{\omega}_i] = 0. \tag{2.73}$$

Similar to d'Alembert's principle, all the reactions disappeared. However, the virtual displacements are replaced by the virtual velocities and the sometimes cumbersome evaluation of the virtual rotations is dropped. Further, Jourdain's principle handles nonlinear and non-holonomic constraints, too, which may appear in controlled vehicle systems.

In the American literature, Jourdain's principle is referred to as Kane's equations and the virtual velocities are denoted as partial velocities (see Kane and Levinson [11]). Applying these principles for the generation of the equations of motion, the reactions have not to be considered at all. Therefore, the principles may be also classified as an analytical approach.

2.3.4 Energy Considerations and Lagrange's Equations

An alternative for the generation of the equations of motion is the analytical method by Lagrange (1788) based on energy considerations. The kinetic energy T of a rigid body reads as

$$T = \frac{1}{2} m \boldsymbol{v}_C^2 + \frac{1}{2} \boldsymbol{\omega} \cdot \boldsymbol{I}_C \cdot \boldsymbol{\omega}, \tag{2.74}$$

where the inertia properties $\{m, \boldsymbol{I}_C\}$ and the velocity properties $\{\boldsymbol{v}_C, \omega\}$ are related to the center of mass. The kinetic energy is composed by the translational and rotational energy of the body; it is a scalar quantity, which may be computed in different frames, too.

The kinetic energy of a multi-body system consisting of the bodies K_i, $i = 1(1)p$ comprises the kinetic energy of all bodies as

$$T = \frac{1}{2} \sum_{i=1}^{p} \left[(\boldsymbol{v}_i^I)^{\mathrm{T}} m_i \, \boldsymbol{v}_i^I + (\boldsymbol{\omega}_i^I)^{\mathrm{T}} \boldsymbol{I}_i^I \boldsymbol{\omega}_i^I \right], \tag{2.75}$$

written consistently in the inertial frame I and related to the center of mass C_i of each body K_i. If the work of the applied forces is independent of the path, then the forces have a potential U and it yields

$$\boldsymbol{f}^{(e)} = -\operatorname{grad} U, \tag{2.76}$$

where U is a scalar function of the position. Forces satisfying 2.76 are called conservative; they do not change the total energy of the system. In contrary, nonconservative forces change the total energy; they are called dissipative since the total energy is decreasing. Conservative forces may be due to gravity, $f_G = mg$, or elasticity, $f_F = -ks$. The corresponding potentials read as

$$U_G = m \, g \, z, \quad U_F = \frac{1}{2} k s^2, \tag{2.77}$$

where

z represents the vertical displacement of the center of mass of a body with mass m in the direction of gravity with acceleration g

s means the displacement of an elastic spring with coefficient k

To the potentials, a constant may be added, that is, the origin of a potential can be arbitrarily chosen. The potential energy U of a multi-body system is given by the sum of the body potentials $U = \sum U_j$. Multi-body systems subject to conservative forces only are called conservative systems. For such systems, it yields the energy conservation law:

$$T + U = T_0 + U_0 = \text{constant}. \tag{2.78}$$

The energy conservation law may be derived from Newton's and Euler's law, that is, it does not contain new information. Its application is advantageous for conservative systems with one degree of freedom to evaluate a relation between the position and velocity variable. If there are two different positions known, the unknown velocity can be found from 2.78.

Based on energy expressions, the equations of motion of multi-body systems may be found, too. This will be shown for multi-body systems with holonomic constraints. In contrary to the synthetic method, the bodies of the system have not to be dismantled, the system is considered as a whole. For this purpose, the generalized coordinates $\boldsymbol{y}$ are defined, and the position and the velocity variables (2.37 through 2.39) are evaluated. As a result, the kinetic energy is available as a function of $y_k(t)$ and $\dot{y}_k(t)$, $k = 1(1)f$,

$$T = T(y_k, \dot{y}_k). \tag{2.79}$$

The applied forces and torques are projected in the direction of the generalized coordinates and composed to the generalized forces

$$q_k = \sum_{i=1}^{p} \left[\left(\frac{\partial \boldsymbol{r}_i^I}{\partial y_k} \right)^{\mathrm{T}} \boldsymbol{f}_i^{(e)I} + \left(\frac{\partial \boldsymbol{s}_i^I}{\partial y_k} \right)^{\mathrm{T}} \boldsymbol{l}_i^{(e)I} \right], \quad \boldsymbol{k} = 1(1)f, \tag{2.80}$$

where the rows of the Jacobian matrices (2.41) are used. The generalized forces may be also found by decomposition of the total work of the applied forces and torques

$$\delta W^e = \sum_{i=1}^{p} \left[\boldsymbol{f}_i^{(e)\mathrm{T}} \delta \boldsymbol{r}_i + \boldsymbol{l}_i^{(e)\mathrm{T}} \delta \boldsymbol{s}_i \right] = \sum_{k=1}^{f} q_k \delta y_k. \tag{2.81}$$

In any case, the reaction forces and torques do not appear.

Now the Lagrangian equations of the second kind read as, see, for example, Magnus and Müller [9],

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial T}{\partial \dot{y}_k} \right) - \frac{\partial T}{\partial y_k} = q_k \,, \quad k = 1(1)f. \tag{2.82}$$

For the evaluation of the equations of motion, two partial and one total differentiations have to be performed with respect to one scalar function $T(y_k, \dot{y}_k)$.

As a result, the minimal number f of equations of motion is found. However, the reaction forces are completely lost and cannot be regained.

For conservative systems, the generalized forces follow immediately from the potential

$$q_k = -\frac{\partial U}{\partial y_k}. \tag{2.83}$$

From 2.82 and 2.83, it remains

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial T}{\partial \dot{y}_k} \right) - \frac{\partial T}{\partial y_k} + \frac{\partial U}{\partial y_k} = 0 \,, \quad k = 1(1)f. \tag{2.84}$$

Introducing the Lagrange function $L = T - U$, also called the kinetic potential, 2.84 is even more simplified and reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{y}_k}\right) - \frac{\partial L}{\partial y_k} = 0\,, \quad L = T - U, \quad k = 1(1)f. \tag{2.85}$$

For some engineering applications, it is advantageous to use surplus coordinates $\bar{y}_j$, $j = 1(1)f + r$, in addition to the f-generalized coordinates y_k, $k = 1(1)f$. Then, there exist r geometric constraints between the surplus coordinates

$$\varphi_n = \varphi_n(\bar{y}_j) = 0, \quad n = 1(1)r. \tag{2.86}$$

The equations of motion are now extended by r Lagrangian multipliers λ_n, $n = 1(1)r$, representing generalized constraint forces

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\bar{y}}_j}\right) - \frac{\partial T}{\partial \bar{y}_j} = \bar{q}_j + \sum_{n=1}^{r} \lambda_n \frac{\partial \varphi_n}{\partial \bar{y}_j}, \quad j = 1(1)f + r. \tag{2.87}$$

Thus, a set of differential-algebraic equations (DAEs) remain.

2.4 EQUATIONS OF MOTION

In Section 2.2, there has been presented two methods for the generation of the equations of motion: the synthetic method by Newton–Euler and the analytic method by d'Alembert, Jourdain, or Lagrange, respectively. The principal steps in the generation process by Newton–Euler and Lagrange are shown in Figure 2.9 as block diagram. Common starting point is the mechanical model of the vehicle composed by the elements of multi-body systems. Common result are the equations of motion; they are identical with both methods if the same generalized coordinates are used. However, the effort is different. During the generation of the equations of motion using Lagrange's equations, there appear terms in $\mathrm{d}/\mathrm{d}t\,(\partial T/\partial \dot{y}_i)$ which are afterward eliminated by $\partial T/\partial y_i$ according to 2.82. This means an useless computational effort which is not required with the Newton–Euler approach, see, e.g., Schiehlen and Eberhard [10]. On the other hand, in the Newton–Euler equations, the reactions have to be eliminated. Thus, both of these approaches have disadvantages, which are avoided by a combination of the Newton–Euler equations with the principles presented in Section 2.3.3.

The resulting equations of motion are always ODEs. However, their form depends on the type of the multi-body system. There are ideal and nonideal systems: the first ones are characterized by applied forces and torques independent from any reaction while the second ones show a dependency. For example, gravitational forces, spring, and damper forces are independent from any reactions while sliding friction forces and slip-dependent contact forces, regularly found with tires in vehicle dynamics, are a function of the normal or reaction forces, respectively.

Within the class of ideal systems, ordinary and general multi-body systems are distinguished. Ordinary multi-body systems are due to holonomic constraints and applied forces depending only on position and velocity quantities; they can be always represented by a system of differential equations of the second order. For non-holonomic constraints and/or general force laws, one gets general multi-body systems.

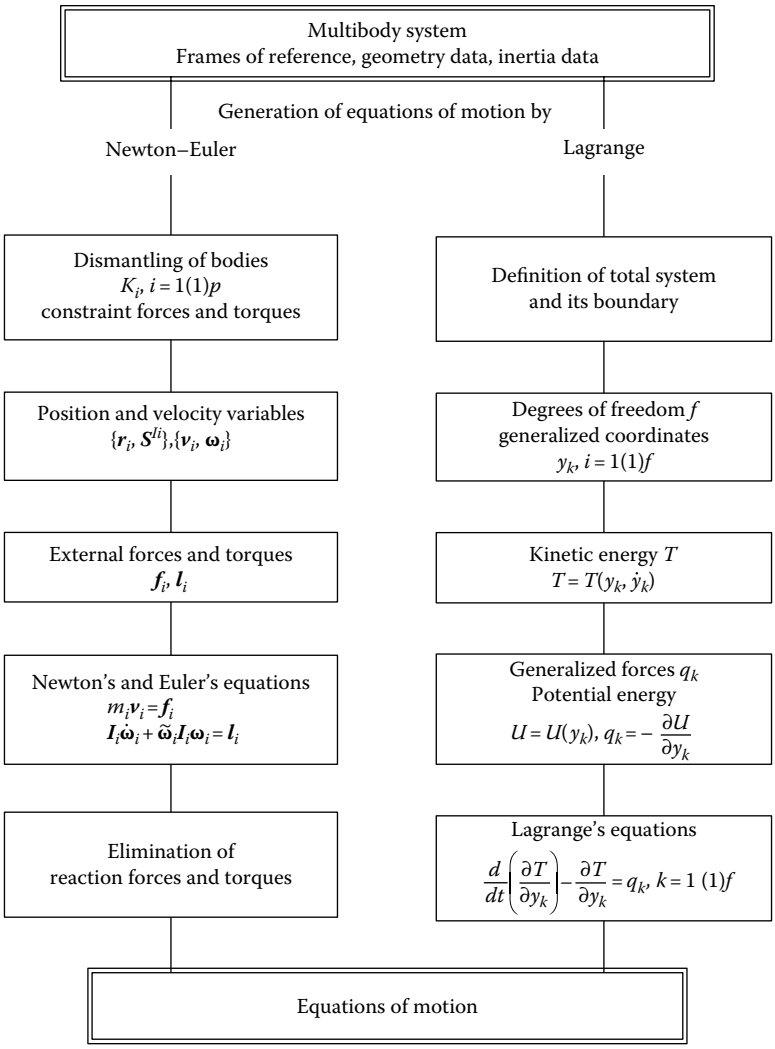


FIGURE 2.9 Generation of equations of motion for multi-body systems by Newton–Euler and Lagrange.

The equations of ordinary multi-body systems read as

$$\boldsymbol{M}(\boldsymbol{y}, t)\,\ddot{\boldsymbol{y}}(t) + \boldsymbol{k}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) = \boldsymbol{q}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t), \tag{2.88}$$

where

$\boldsymbol{y}$ is the $f \times 1$ position vector of the generalized coordinates
$\boldsymbol{M}$ is the $f \times f$ symmetric inertia matrix
$\boldsymbol{k}$ is a $f \times 1$ vector of generalized gyroscopic forces including the Coriolis and centrifugal forces as well as the gyroscopic torques
$f \times 1$ vector $\boldsymbol{q}$ represents generalized applied forces

The equations of motion resulting from the analytical method have always the form (2.88) while the synthetic method often requires some calculations to get a symmetric inertia matrix.

In vehicle dynamics, the deviations $\tilde{\boldsymbol{y}}(t)$ from a reference motion $\boldsymbol{y} = \boldsymbol{y}_R(t)$ are often small:

$$\boldsymbol{y}(t) = \boldsymbol{y}_R(t) + \tilde{\boldsymbol{y}}(t). \tag{2.89}$$

Then, it follows by a Taylor series expansion under assumption of differentiable vector functions, and skipping of the second- and higher-order terms from 2.88 the linearized equations of motion, see also Müller and Schiehlen [12],

$$\boldsymbol{M}(t)\ddot{\tilde{\boldsymbol{y}}}(t)+\boldsymbol{P}(t)\dot{\tilde{\boldsymbol{y}}}(t)+\boldsymbol{Q}(t)\tilde{\boldsymbol{y}}(t)=\boldsymbol{h}(t), \tag{2.90}$$

where

$\boldsymbol{M}(t)$ is the symmetric, positive definite inertia matrix
$\boldsymbol{P}(t)$ and $\boldsymbol{Q}(t)$ characterize the velocity- and position-dependent forces
the vector $\boldsymbol{h}(t)$ represents the external excitations

If all these matrices are time-invariant and subdivided in a symmetrical and skew-symmetrical part, then the equations of motion of a linear ordinary and time-invariant multi-body system are found reading as

$$\boldsymbol{M}\ddot{\boldsymbol{y}}(t)+(\boldsymbol{D}+\boldsymbol{G})\dot{\boldsymbol{y}}(t)+(\boldsymbol{K}+\boldsymbol{N})\,\boldsymbol{y}(t)=\boldsymbol{h}(t), \tag{2.91}$$

where $\tilde{\boldsymbol{y}}$ used in 2.90 is simply replaced by $\boldsymbol{y}$ and the $f\times f$ matrices have the properties

$$\boldsymbol{M}=\boldsymbol{M}^{\mathrm{T}}>\boldsymbol{0}\,,\quad \boldsymbol{D}=\boldsymbol{D}^{\mathrm{T}},\quad \boldsymbol{G}=-\boldsymbol{G}^{\mathrm{T}},\quad \boldsymbol{K}=\boldsymbol{K}^{\mathrm{T}},\quad \boldsymbol{N}=-\boldsymbol{N}^{\mathrm{T}}. \tag{2.92}$$

These matrices have a physical meaning which can be identified after pre-multiplication of 2.92 from the left by $\dot{\boldsymbol{y}}^{\mathrm{T}}$ resulting in the total time derivative of an energy expression:

$$\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{M}\ddot{\boldsymbol{y}}}+\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{D}\dot{\boldsymbol{y}}}+\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{G}\dot{\boldsymbol{y}}}+\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{y}}+\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{y}}=\underbrace{\dot{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{h}}, \tag{2.93}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}T+2R+0+\frac{\mathrm{d}}{\mathrm{d}t}U+2S=P. \tag{2.94}$$

The inertia matrix $\boldsymbol{M}$ determines the kinetic energy $T=\frac{1}{2}\dot{\boldsymbol{y}}^{T}\boldsymbol{M}\,\dot{\boldsymbol{y}}$ and, therefore, the inertia forces, from $T>0$ it follows again the positive definiteness of the inertia matrix. The damping matrix $\boldsymbol{D}$ defines via Rayleigh's dissipation function $R=\frac{1}{2}\dot{\boldsymbol{y}}^{T}\boldsymbol{D}\,\dot{\boldsymbol{y}}$ the damping forces while the gyro matrix $\boldsymbol{G}$ describes the gyroscopic forces, which do not change the total energy of the system. The stiffness matrix determines the potential energy $U=\frac{1}{2}\boldsymbol{y}^{T}\boldsymbol{K}\,\boldsymbol{y}$ and, therefore, the conservative position forces while the matrix $\boldsymbol{N}$ identifies the circulatory forces also known as nonconservative position forces. Furthermore, P describes the power of the external excitation forces. For $\boldsymbol{D}=0$, $\boldsymbol{N}=0$, and $\boldsymbol{h}=0$, the multi-body system is conservative, that is, the total energy is constant for all motions,

$$T+U=\text{constant}. \tag{2.95}$$

The matrix properties (2.92) allow often to check the equations of motion with respect to the physical phenomena involved.

In general, multi-body systems exist a broader variety of standard representations, which will not be discussed in detail; anyway, the reader can refer to [10] for additional information.

2.5 FORMALISMS FOR MULTI-BODY SYSTEMS

The generation of equations of motion for large multi-body systems is a nontrivial task requiring numerous steps during the evaluation of the fundamental relations. Beginning with the space age in the middle of the 1960s, the generation of equations of motion was more formalized. The resulting formalisms were used for the development of computer codes for multi-body systems; they are the basis of computational multi-body dynamics. Twenty-five years later, in 1990, there were known 20 formalisms described in the *Multibody System Handbook* [7]. Many of them are used today.

Multi-body system formalisms are based on Newton–Euler equations or Lagrange's equations, respectively, as described in Sections 2.3.2 and 2.3.3. Regarding the computational procedure, numerical and symbolical formalisms are distinguished (Figure 2.10). Numerical formalisms supply the elements of the system matrices as numbers for linear time-invariant multi-body systems (2.91). In the case of linear time-variant systems (2.90) and nonlinear systems (2.88), a numerical formalism provides the numbers in the equations of motion necessary for each time step required by the simulation program. In contrary, symbolical formalisms generate the equations of motion only once with the computer how it is done with paper and pencil. The advantage is that variations of the system parameters and, for time-variant systems, the current time have to be inserted in the symbolical equations of motion only (see Figure 2.10). Symbolical formalisms are especially helpful for optimizations and control design.

Furthermore, non-recursive and recursive formalisms are distinguished. Recursive formalisms make use of special topology properties of multi-body systems.

2.5.1 Non-Recursive Formalisms

First of all, the symbolical formalism NEWEUL is presented. NEWEUL is a research software based on the Newton–Euler equations and the principles of d'Alembert and Jourdain; it was developed by Kreuzer at the University of Stuttgart, and recently revised in Matlab as Neweul-M power 2 (Neweul M squared), see [13]. NEWEUL generates equations of motion in minimal form (2.88) and (2.90), respectively, which may be solved by any integration code for ODEs. The formalisms comprise five steps, which may be evaluated by hand for smaller multi-body systems, too.

Step 1: System specification and input data.

At first the multi-body system is defined and treated as whole. The number of degrees of freedom is determined and the generalized coordinates y_k, $k = 1(1)f$ are chosen. The inertial frame I and

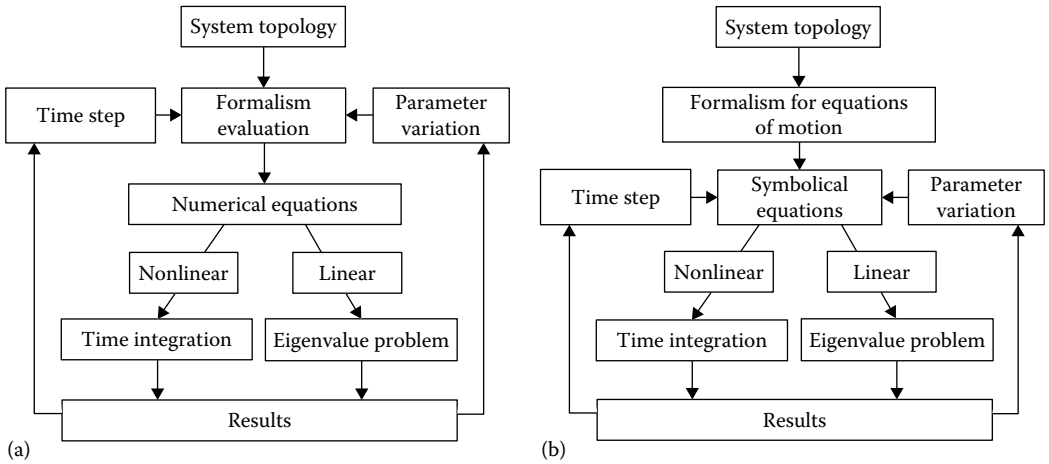


FIGURE 2.10 Flow diagrams of multi-body dynamical analysis by (a) numerical formalisms and (b) symbolical formalisms.

the body-fixed frames i are chosen. Each body K_i, $i = 1(1)p$ is dismantled, and the corresponding inertia parameters $\{m_i, \boldsymbol{I}_i\}$, the position variables $\left\{ \boldsymbol{r}_i^I, \boldsymbol{S}^{Ii} \right\}$, as well as the applied forces and torques $\left\{ \boldsymbol{f}_i^{(e)I}, \boldsymbol{l}_i^{(e)I} \right\}$ are specified. The quantities $\boldsymbol{I}_i^I$ and $\boldsymbol{l}_i^{(e)I}$ are related to the corresponding center of mass C_i. The input data read as

$$f, \boldsymbol{y} = [y_1, \ldots, y_j, \ldots, y_f]^{\mathrm{T}}, \tag{2.96}$$

$$p, \ \left\{ m_i, \boldsymbol{I}_i^i \right\}, \ \left\{ \boldsymbol{r}_i^I, \boldsymbol{S}^{Ii} \right\}, \ \left\{ \boldsymbol{f}_i^{(e)I}, \boldsymbol{l}_i^{(e)I} \right\}, \quad i = 1(1)p, \tag{2.97}$$

$$\boldsymbol{r}_i^I = \left[r_{i1} \ \ r_{i2} \ \ r_{i3} \right]^{\mathrm{T}}, \quad \boldsymbol{S}^{Ii} = \boldsymbol{S}^{Ii}\left(\alpha_i, \beta_i, \gamma_i \right) \equiv \boldsymbol{S}_i, \quad i = 1(1)p. \tag{2.98}$$

The rotation matrices are described by three angles, for example, the Cardan angles α_i, β_i, γ_i.

Step 2: Element consideration, local equations.

At first, the elements of the inertia tensors are computed in the inertial frame I by the transformation

$$\boldsymbol{I}_i^I = \boldsymbol{S}^{Ii} \boldsymbol{I}_i^i \boldsymbol{S}^{iI}. \tag{2.99}$$

Now, there are all quantities available in the inertial frame I which is the only frame further used. Thus, the right upper index is no longer required and just skipped. The local equations of motion for each body K_i read as

$$m_i \dot{\boldsymbol{v}}_i = \boldsymbol{f}_i^{(e)} + \boldsymbol{f}_i^{(r)}, \quad i = 1(1)p, \tag{2.100}$$

$$\boldsymbol{I}_i \dot{\boldsymbol{\omega}}_i + \tilde{\boldsymbol{\omega}}_i \boldsymbol{I}_i \boldsymbol{\omega}_i = \boldsymbol{l}_i^{(e)} + \boldsymbol{l}_i^{(r)}, \quad i = 1(1)p, \tag{2.101}$$

where the external forces and torques are subdivided in the known applied forces and torques and the unknown reactions. The reactions are eliminated later and, therefore, they are not to be specified.

Step 3: Relation between local and global quantities.

The relation between the position (2.98) of a single body K_i and the generalized coordinates (2.96) is given by the holonomic, and in general, also by rheonomic constraints. These relations are known from the input data as

$$\boldsymbol{r}_i = \boldsymbol{r}_i(\boldsymbol{y}, t), \quad \boldsymbol{S}_i = \boldsymbol{S}_i(\boldsymbol{y}, t), \quad i = 1(1)p, \tag{2.102}$$

and the corresponding velocities $\{\boldsymbol{v}_i, \omega_i\}$ and accelerations $\{\boldsymbol{a}_i, \alpha_i\}$ are computed as

$$\boldsymbol{v}_i = \dot{\boldsymbol{r}}_i = \frac{\partial \boldsymbol{r}_i}{\partial \boldsymbol{y}^{\mathrm{T}}} \dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{r}_i}{\partial t} = \boldsymbol{J}_{Ti}(\boldsymbol{y}, t) \dot{\boldsymbol{y}} + \overline{\boldsymbol{v}}_i(\boldsymbol{y}, t), \tag{2.103}$$

$$\boldsymbol{\omega}_i = \dot{\boldsymbol{s}}_i = \frac{\partial \boldsymbol{s}_i}{\partial \boldsymbol{y}^{\mathrm{T}}} \dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{s}_i}{\partial t} = \boldsymbol{J}_{Ri}(\boldsymbol{y}, t) \dot{\boldsymbol{y}} + \overline{\boldsymbol{\omega}}_i(\boldsymbol{y}, t), \tag{2.104}$$

$$\boldsymbol{a}_i = \dot{\boldsymbol{v}}_i = \boldsymbol{J}_{Ti}(\boldsymbol{y}, t) \ddot{\boldsymbol{y}} + \frac{\partial \boldsymbol{v}_i}{\partial \boldsymbol{y}^{\mathrm{T}}} \dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{v}_i}{\partial t}, \tag{2.105}$$

$$\boldsymbol{a}_i = \dot{\boldsymbol{\omega}}_i = \boldsymbol{J}_{Ri}(\boldsymbol{y},t)\,\ddot{\boldsymbol{y}} + \frac{\partial \boldsymbol{\omega}_i}{\partial \boldsymbol{y}^{\mathrm{T}}}\,\dot{\boldsymbol{y}} + \frac{\partial \boldsymbol{\omega}_i}{\partial t}. \tag{2.106}$$

For scleronomic constraints, the partial time derivatives are disappearing.

The $3 \times f$ Jacobian matrices $\boldsymbol{J}_{Ti}$, $\boldsymbol{J}_{Ri}$ of translation and rotation, respectively, present the relation between the local and global coordinates as shown by the virtual motion:

$$\delta \boldsymbol{r}_i = \boldsymbol{J}_{Ti}\,\delta \boldsymbol{y}, \quad \delta s_i = \boldsymbol{J}_{Ri}\,\delta \boldsymbol{y}, \quad i = 1(1)p. \tag{2.107}$$

These matrices are available from kinematics from the relations (2.103) and (2.104).

After these preparatory computations, the local equations (2.100) and (2.101) of body K_i of the multi-body system are evaluated as functions of the generalized coordinates and their derivatives.

Step 4: System consideration, global equations.

At first the local equ ations, depending on the generalized coordinates, are composed in global matrices and vectors. For this purpose, the $6p \times 6p$ diagonal matrix $\overline{\overline{\boldsymbol{M}}}$ of the inertia quantities is introduced

$$\overline{\overline{\boldsymbol{M}}} = \mathrm{diag}\,[m_1\,\boldsymbol{E}, m_2\,\boldsymbol{E}, \ldots, m_p\,\boldsymbol{E}, \boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_p], \tag{2.108}$$

where $\boldsymbol{E}$ means the 3×3 unit matrix. Further, the $6p \times 1$ vectors $\overline{\boldsymbol{q}}^{(e)}$, $\overline{\boldsymbol{q}}^{(r)}$, $\boldsymbol{k}$ are used to summarize all applied reaction, Coriolis and gyro forces, and torques. These three vectors are defined as follows:

$$\overline{\boldsymbol{q}} = \left[\boldsymbol{f}_1^{\mathrm{T}}, \ldots, \boldsymbol{f}_p^{\mathrm{T}}, \boldsymbol{l}_1^{\mathrm{T}}, \ldots, \boldsymbol{l}_p^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.109}$$

And finally, the global $6p \times f$ Jacobian matrix is introduced as

$$\overline{\boldsymbol{J}} = [\boldsymbol{J}_{T1}^{\mathrm{T}}, \ldots, \boldsymbol{J}_{Tp}^{\mathrm{T}}, \boldsymbol{J}_{R1}^{\mathrm{T}}, \ldots, \boldsymbol{J}_{Rp}^{\mathrm{T}}]^{\mathrm{T}}. \tag{2.110}$$

The global Newton–Euler equations are now represented as one $6p \times 1$ vector equation:

$$\overline{\overline{\boldsymbol{M}}}\,\overline{\boldsymbol{J}}\,\ddot{\boldsymbol{y}} + \overline{\boldsymbol{k}}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) = \overline{\boldsymbol{q}}^{(e)}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) + \overline{\boldsymbol{q}}^{(r)}. \tag{2.111}$$

These $6p$ equations are reduced to the minimal number of f ODEs by left pre-multiplication with the transposed $f \times 6p$ Jacobian matrix $\overline{\boldsymbol{J}}^{\mathrm{T}}$,

$$\overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}\,\overline{\boldsymbol{J}}\ddot{\boldsymbol{y}} + \overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\boldsymbol{k}}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) = \overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\boldsymbol{q}}^{(e)}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t), \tag{2.112}$$

where the term $\overline{\boldsymbol{J}}^{\mathrm{T}}\boldsymbol{q}(r)$ is vanishing due to 2.62.

Summarizing the matrix products, one gets the equations of motion with symmetric inertia matrix

$$\boldsymbol{M}(\boldsymbol{y}, t)\ddot{\boldsymbol{y}}(t) + \boldsymbol{k}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) = \boldsymbol{q}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t),$$

$$\boldsymbol{M} = \boldsymbol{M}^{\mathrm{T}} = \overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}\,\overline{\boldsymbol{J}}, \quad \boldsymbol{k} = \overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\boldsymbol{k}}, \quad \boldsymbol{q} = \overline{\boldsymbol{J}}^{\mathrm{T}}\overline{\boldsymbol{q}}^{(e)} \tag{2.113}$$

in complete agreement with 2.88.

Optional step 5: Computation of the reaction forces and torques.

In contrary to Lagrange's approach, the reactions can be regained if the global distribution matrix $\bar{\boldsymbol{Q}}$ according to 2.62 and the generalized constraint forces $\boldsymbol{g}$ from 2.63 are employed. Then, it yields

$$\bar{\boldsymbol{q}}^{(r)} = \bar{\boldsymbol{Q}}\boldsymbol{g}. \tag{2.114}$$

Due to the orthogonality condition (2.68), the left pre-multiplication of 2.111 with $\bar{\boldsymbol{Q}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}^{-1}$ results in linear, completely algebraic reaction equations

$$\hat{\boldsymbol{N}}(\boldsymbol{y},t)\,\boldsymbol{g}(t) + \hat{\boldsymbol{q}}(\boldsymbol{y},\dot{\boldsymbol{y}},t) = \hat{\boldsymbol{k}}(\boldsymbol{y},\dot{\boldsymbol{y}},t),$$

$$\hat{\boldsymbol{N}} = \hat{\boldsymbol{N}}^{\mathrm{T}} = \bar{\boldsymbol{Q}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}^{-1}\bar{\boldsymbol{Q}},\ \hat{\boldsymbol{q}} = \bar{\boldsymbol{Q}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}^{-1}\bar{\boldsymbol{q}}^{(e)},\ \hat{\boldsymbol{k}} = \bar{\boldsymbol{Q}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}^{-1}\bar{\boldsymbol{k}}, \tag{2.115}$$

where $\hat{\boldsymbol{N}}$is the symmetric, generally positive definite $q \times q$ reaction matrix while the $q \times 1$ vectors $\hat{\boldsymbol{q}}$, $\hat{\boldsymbol{k}}$ show the influence of the applied and gyroscopic forces and torques on the reactions. On the other hand, by elimination of $\ddot{\boldsymbol{y}}$ in 2.111, one gets

$$\bar{\boldsymbol{q}}^{(r)} = \bar{\boldsymbol{Q}}\boldsymbol{g} = [\boldsymbol{E} - \overline{\overline{\boldsymbol{M}}}\ \bar{\boldsymbol{J}}(\bar{\boldsymbol{J}}^{\mathrm{T}}\overline{\overline{\boldsymbol{M}}}\,\bar{\boldsymbol{J}})^{-1}\bar{\boldsymbol{J}}^{\mathrm{T}}](\bar{\boldsymbol{k}} - \bar{\boldsymbol{q}}^{(e)}). \tag{2.116}$$

Then, by partitioning of 2.116, only some of the constraint forces and torques may be computed.

In vehicle dynamics, often contact forces are found resulting in nonideal multi-body systems. Then, the f equations of motion (2.113) have to be solved simultaneously with the q equations of reaction (2.115).

The presented five steps show that the NEWEUL formalism is based on the Newton–Euler equations; however, they are supplemented by typical features of the analytical method like generalized coordinates and generalized reactions. The required computations include summation, multiplication, and differentiation of vectors and matrices, simplification of trigonometric expressions, and linearization of expressions. These computations are symbolically performed in NEWEUL. But the underlying formalism may also be executed by any formula manipulation software, for example, by MAPLE. More details on NEWEUL are found on the web: www.itm.uni-stuttgart.de/research/neweul/neweulm2_en.php

The most widely used numerical formalism in vehicle engineering is MSC.ADAMS. The software MSC.ADAMS generate numerical solutions of the original nonreduced Newton–Euler equations (2.55) and (2.56) rewritten as

$$\overline{\overline{\boldsymbol{M}}}\ \ddot{\boldsymbol{x}} + \boldsymbol{k}\left(\boldsymbol{x},\dot{\boldsymbol{x}},t\right) = \bar{\boldsymbol{q}}^{(e)}\left(\boldsymbol{x},\dot{\boldsymbol{x}},t\right) + \bar{\boldsymbol{Q}}\boldsymbol{g},\quad \bar{\boldsymbol{Q}}^{\mathrm{T}} = -\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{x}^{\mathrm{T}}} = -\boldsymbol{\Phi}_x, \tag{2.117}$$

where the $6p \times 1$ vector $\boldsymbol{x}(t)$ is composed by Cartesian coordinates and the $6p \times q$ distribution matrix $\bar{\boldsymbol{Q}}$ of the reactions is found from the implicit constraints (2.33).

After two total time derivatives of 2.33, a set of $6p + q$ linear equations remains for the unknowns $\ddot{\boldsymbol{x}}$and $\boldsymbol{g}$,

$$\begin{bmatrix} \overline{\overline{\boldsymbol{M}}} & \boldsymbol{\Phi}_x^{\mathrm{T}} \\ \boldsymbol{\Phi}_x & \boldsymbol{0} \end{bmatrix}\begin{bmatrix} \ddot{\boldsymbol{x}} \\ \boldsymbol{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}^{(e)} - \boldsymbol{k} \\ -\boldsymbol{\Phi}_t - \dot{\boldsymbol{\Phi}}_x\dot{\boldsymbol{x}} \end{bmatrix}. \tag{2.118}$$

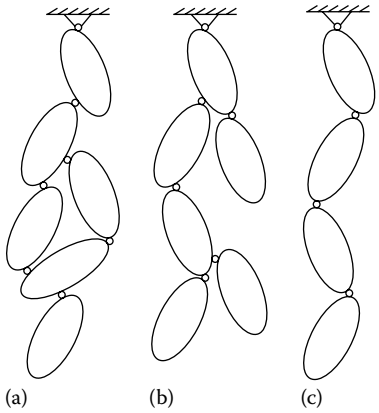


FIGURE 2.11 Topology of multi-body systems. (a) Loop, (b) tree, (c) chain.

Here, $\overline{\boldsymbol{M}}$ is a block diagonal $6\,p \times 6\,p$ matrix which allows the application of sparse matrix techniques. Equation 2.118 represents DAEs, which are solved within the MSC.ADAMS software with special integration codes (see also Chapter 3). As with all commercial codes, a graphical interface is used, and mathematical equations are not to be handled. For more details, search for ADAMS under www.mscsoftware.com

2.5.2 Recursive Formalisms

For time integration of holonomic systems, the mass matrix in 2.88 or 2.113, respectively, has to be inverted what is numerically costly for systems with many degrees of freedom

$$\ddot{\boldsymbol{y}}(t) = \boldsymbol{M}^{-1}(\boldsymbol{y},t)\left[\boldsymbol{q}(\boldsymbol{y},\dot{\boldsymbol{y}},t) - \boldsymbol{k}(\boldsymbol{y},\dot{\boldsymbol{y}},t)\right]. \tag{2.119}$$

Recursive algorithms avoid this matrix inversion. The fundamental requirement, however, is a chain or tree topology of the multi-body system as shown in Figure 2.11. Loop topologies are not included. Contributions on recursive algorithms are due, for example, to Hollerbach [14], Bae and Haug [15], Brandl et al. [16], and Schiehlen [17].

2.5.2.1 Kinematics

Recursive kinematics uses the relative motion between two neighboring bodies and the related constraints as shown in Figure 2.12. The absolute translational and rotational velocity vector $\boldsymbol{w}_i$ of body i, also denoted as twist, is related to the absolute velocity vector w_{i-1} of body $(i-1)$ and the generalized coordinates $\boldsymbol{y}_i$ of the joint i between this two bodies. It yields

$$\underbrace{\begin{bmatrix}\boldsymbol{v}_{Oi}\\ \boldsymbol{\omega}_i\end{bmatrix}}_{\boldsymbol{w}_i} = \underbrace{\boldsymbol{S}^{i,i-1}\begin{bmatrix}\boldsymbol{E} & -\tilde{\boldsymbol{r}}_{Oi-1,Oi}\\ \boldsymbol{0} & \boldsymbol{E}\end{bmatrix}}_{\boldsymbol{C}_i}\underbrace{\begin{bmatrix}\boldsymbol{v}_{Oi-1}\\ \boldsymbol{\omega}_{i-1}\end{bmatrix}}_{\boldsymbol{w}_{i-1}} + \underbrace{\boldsymbol{S}^{i,i-1}\begin{bmatrix}\boldsymbol{J}_{Ti}\\ \boldsymbol{J}_{Ri}\end{bmatrix}}_{\boldsymbol{J}_i}\dot{\boldsymbol{y}}_i. \tag{2.120}$$

Using the fundamentals of relative motion of rigid bodies, it remains for the absolute acceleration

$$\boldsymbol{b}_i = \boldsymbol{C}_i\,\boldsymbol{b}_{i-1} + \boldsymbol{J}_i\;\ddot{\boldsymbol{y}}_i + \beta_i(\dot{\boldsymbol{y}}_i, \boldsymbol{w}_{i-1}), \tag{2.121}$$

where the vector $\boldsymbol{b}_i$ summarizes the translational and rotational accelerations of body i as well.

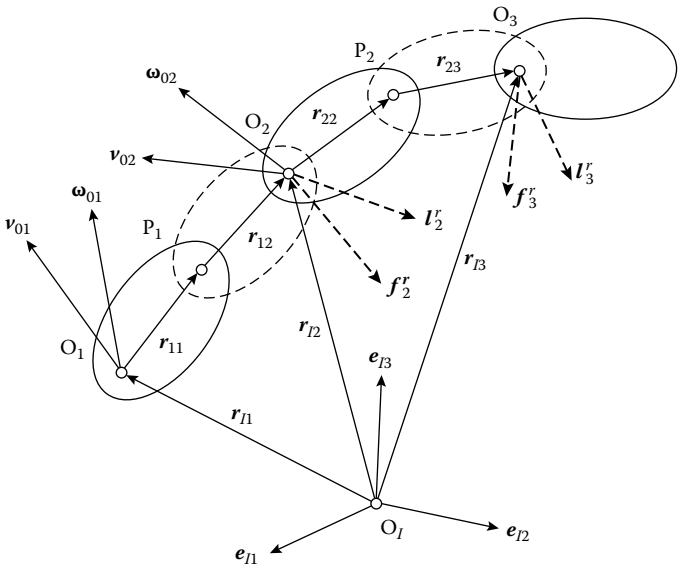


FIGURE 2.12 Three-body system with two joints.

For the total system, one gets for the absolute acceleration in matrix notation

$$\boldsymbol{b} = \boldsymbol{C}\ \boldsymbol{b} + \boldsymbol{J}\,\ddot{\boldsymbol{y}} + \beta, \tag{2.122}$$

where the geometry matrix $\boldsymbol{C}$ is a lower block-sub-diagonal matrix and the Jacobian $\boldsymbol{J}$ is a block-diagonal matrix as follows:

$$\boldsymbol{C} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \boldsymbol{C}_2 & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{C}_p & 0 \end{bmatrix}, \quad \boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{J}_2 & \boldsymbol{0} & \cdots & 0 \\ 0 & 0 & \boldsymbol{J}_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \boldsymbol{J}_p \end{bmatrix}. \tag{2.123}$$

From 2.122, it follows the non-recursive form of the absolute accelerations as

$$\boldsymbol{b} = \left(\boldsymbol{E} - \boldsymbol{C}\right)^{-1} \boldsymbol{J}\,\ddot{\boldsymbol{y}} + \bar{\beta}, \tag{2.124}$$

where the global Jacobian matrix $\bar{\boldsymbol{J}}$ is found again (see 2.67 or 2.110)

$$\bar{\boldsymbol{J}} = \left(\boldsymbol{E} - \boldsymbol{C}\right)^{-1} \boldsymbol{J} = \begin{bmatrix} \boldsymbol{J}_1 & 0 & 0 & \cdots & 0 \\ \boldsymbol{C}_2\boldsymbol{J}_1 & \boldsymbol{J}_2 & 0 & \cdots & 0 \\ \boldsymbol{C}_3\boldsymbol{C}_2\boldsymbol{J}_1 & \boldsymbol{C}_3\boldsymbol{J}_2 & \boldsymbol{J}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & \boldsymbol{J}_p \end{bmatrix}. \tag{2.125}$$

Due to the chain topology, the global Jacobian matrix is a lower triangular matrix.

2.5.2.2 Newton–Euler Equations

Newton's and Euler's equations are now written for body i in its body fixed frame at the joint position O_i using the absolute accelerations and the external forces $\boldsymbol{q}_i$ acting on the body with holonomic constraints:

$$\underbrace{\begin{bmatrix} m_i\boldsymbol{E} & m_i\tilde{\boldsymbol{r}}_{OiCi}^T \\ m_i\tilde{\boldsymbol{r}}_{OiCi} & \boldsymbol{I}_{Oi} \end{bmatrix}}_{\boldsymbol{M}_i = \text{constant}} \underbrace{\begin{bmatrix} \boldsymbol{a}_{Oi} \\ \boldsymbol{\alpha}_i \end{bmatrix}}_{\boldsymbol{b}_i} + \underbrace{\begin{bmatrix} m_i\tilde{\boldsymbol{\omega}}_i\tilde{\boldsymbol{\omega}}_i\boldsymbol{r}_{OiCi} \\ \tilde{\boldsymbol{\omega}}_i\boldsymbol{I}_{Oi}\boldsymbol{\omega}_i \end{bmatrix}}_{\boldsymbol{k}_i} = \underbrace{\begin{bmatrix} \boldsymbol{f}_i \\ \boldsymbol{l}_{Oi} \end{bmatrix}}_{\boldsymbol{q}_i}. \tag{2.126}$$

Moreover, the external forces are composed of applied forces $\boldsymbol{q}_i^{(e)}$ and constraints forces $q_i^{(r)}$ where the generalized constraint forces of the joint i and joint $(i+1)$ appear:

$$\boldsymbol{q}_i = \boldsymbol{q}_i^{(e)} + \boldsymbol{q}_i^{(r)}, \quad \boldsymbol{q}_i^{(r)} = \boldsymbol{Q}_i\boldsymbol{g}_i - \boldsymbol{C}_{i+1}^{\mathrm{T}}\boldsymbol{Q}_{i+1}\boldsymbol{g}_{i+1}. \tag{2.127}$$

2.5.2.3 Equations of Motion

For the total system, a set of 18 scalar equations remains from 2.122, 2.126, and 2.127

$$\boldsymbol{b} = \bar{\boldsymbol{J}}\ \ddot{\boldsymbol{y}} + \bar{\beta}, \tag{2.128}$$

$$\bar{\bar{\boldsymbol{M}}}\boldsymbol{b} + \bar{\boldsymbol{k}} = \boldsymbol{q}^{(e)} + \boldsymbol{q}^{(r)}, \tag{2.129}$$

$$\boldsymbol{q}^{(r)} = \left(\boldsymbol{E} - \boldsymbol{C}\right)^{\mathrm{T}}\ \boldsymbol{Q}\,\boldsymbol{g} = \bar{\boldsymbol{Q}}\,\boldsymbol{g} \tag{2.130}$$

with 18 unknowns in the vectors $\boldsymbol{b}$, $\boldsymbol{y}$, $\boldsymbol{q}^{(r)}$, $\boldsymbol{g}$.

Now, 2.128 and 2.130 are inserted in 2.129 and the global orthogonality $\bar{\boldsymbol{J}}^T\,\bar{\boldsymbol{Q}} = \boldsymbol{0}$ is used again, resulting in the standard form (2.113) of the equations of motion. The mass matrix is completely full, again, and the vector $\boldsymbol{k}$ depends not only on the generalized velocities but also on the absolute velocities:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{J}_1^t\left(\boldsymbol{M}_1 + \boldsymbol{C}_2^t\left(\boldsymbol{M}_2 + \boldsymbol{C}_3^t\boldsymbol{M}_3\boldsymbol{C}_3\right)\boldsymbol{C}_2\right)\boldsymbol{J}_1 & \boldsymbol{J}_1^t\boldsymbol{C}_2^t\left(\boldsymbol{M}_2 + \boldsymbol{C}_3^t\boldsymbol{M}_3\boldsymbol{C}_3\right)\boldsymbol{J}_2 & \boldsymbol{J}_1^t\boldsymbol{C}_2^t\boldsymbol{C}_3^t\boldsymbol{M}_3\boldsymbol{J}_3 \\ \boldsymbol{J}_2^t\left(\boldsymbol{M}_2 + \boldsymbol{C}_3^t\boldsymbol{M}_3\boldsymbol{C}_3\right)\boldsymbol{C}_2\boldsymbol{J}_1 & \boldsymbol{J}_2^t\left(\boldsymbol{M}_2 + \boldsymbol{C}_3^t\boldsymbol{M}_3\boldsymbol{C}_3\right)\boldsymbol{J}_2 & \boldsymbol{J}_2^t\boldsymbol{C}_3^t\boldsymbol{J}_3 \\ \boldsymbol{J}_3\boldsymbol{M}_3\boldsymbol{C}_3\boldsymbol{C}_2\boldsymbol{J}_1 & \boldsymbol{J}_3^t\boldsymbol{M}_3\boldsymbol{C}_3\boldsymbol{J}_2 & \boldsymbol{J}_3^t\boldsymbol{M}_3\boldsymbol{J}_3 \end{bmatrix}, \tag{2.131}$$

$$\boldsymbol{k} = \boldsymbol{k}\left(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{w}\right). \tag{2.132}$$

However, the mass matrix shows now a characteristic structure, which can be used for a Gauss transformation.

2.5.2.4 Recursion

There are three steps required to obtain the generalized accelerations:

1. Forward recursion to get the absolute motion starting with $i = 1$.
2. Backward recursion using a Gauss transformation starting with $i = p$. As a result, the system

$$\hat{\boldsymbol{M}}\,\ddot{\boldsymbol{y}} + \hat{\boldsymbol{k}} = \hat{\boldsymbol{q}} \tag{2.133}$$

is obtained where $\hat{\boldsymbol{M}}$ is a lower triangular matrix

$$\overset{\Delta}{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{J}_1^t \tilde{\boldsymbol{M}}_1 \boldsymbol{J}_1 & 0 & 0 \\ \boldsymbol{J}_2^t \tilde{\boldsymbol{M}}_2 \boldsymbol{C}_2 \boldsymbol{J}_1 & \boldsymbol{J}_2^t \tilde{\boldsymbol{M}}_2 \boldsymbol{J}_2 & 0 \\ \boldsymbol{J}_3^t \tilde{\boldsymbol{M}}_3 \boldsymbol{C}_3 \boldsymbol{C}_2 \boldsymbol{J}_1 & \boldsymbol{J}_3^t \tilde{\boldsymbol{M}}_3 \boldsymbol{C}_3 \boldsymbol{J}_2 & \boldsymbol{J}_3^t \tilde{\boldsymbol{M}}_3 \boldsymbol{J}_3 \end{bmatrix}, \tag{2.134}$$

$$\tilde{\boldsymbol{M}}_{i-1} = \boldsymbol{M}_{i-1} + \boldsymbol{C}_i^t \left(\tilde{\boldsymbol{M}}_i - \tilde{\boldsymbol{M}}_i \boldsymbol{J}_i \left(\boldsymbol{J}_i^t \tilde{\boldsymbol{M}}_i \boldsymbol{J}_i \right)^{-1} \boldsymbol{J}_i^t \tilde{\boldsymbol{M}}_i \right) \boldsymbol{C}_i \tag{2.135}$$

the block elements of which follow from the recursion formula (2.135).

3. Forward recursion for the generalized accelerations starting with $i = 1$.

The recursion requires some computational overhead. Therefore, the recursive algorithms are more efficient than the matrix inversion for more than $p = 8$–10 bodies.

There are also some extensions of the recursive approach to loop topologies (see Bae and Haug [18] and Saha and Schiehlen [19]). Further, it is mentioned that there are two commercial codes based on recursive formalisms: SIMPACK and RecurDyn. For more details, see www.simpack.de and www.functionbay.co.kr

REFERENCES

1. Schiehlen, W.: Multibody system dynamics: Roots and perspectives. *Multibody System Dynamics* 1 (1997), 149–188.
2. Popp, K. and Schiehlen, W.: *Fahrzeugdynamik*. Teubner, Stuttgart, Germany, 1993. English edition; Ground Vehicle Dynamics. Springer, Berlin, 2010. Chinese translation; Beijing, China Communication Press, 2012.
3. Ammon, D.: *Modellbildung und Systementwicklung in der Fahrzeugdynamik*. Teubner, Stuttgart, Germany, 1997.
4. Dukkipati, R.V.: *Vehicle Dynamics*. CRC Press, Boca Raton, FL, 2000.
5. Wong, J.Y.: *Theory of Ground Vehicles*. Wiley, New York, 2001.
6. Berg, M. and Stensson Trigell, A. (eds.): 21th Int. Symposium on Dynamics of Roads and Tracks. Vehicle System Dynamics 48 (2010), Supplement 1.
7. Schiehlen, W. (ed.). *Multibody System Handbook*. Springer, Berlin, Germany, 1990.
8. Kortüm, W. and Sharp, R.: *Multibody Computer Codes in Vehicle System Dynamics*. Swets & Zeitlinger, Lisse, the Netherlands, 1993.
9. Magnus, K. and Müller, H.H.: *Grundlagen der Technischen Mechanik*. Teubner, Wiesbaden, Germany, 1990.
10. Schiehlen, W. and Eberhard, P.: *Technische Dynamik*. Teubner, Wiesbaden, Germany, 2004.
11. Kane. T.R. and Levinson, D.A.: *Dynamics: Theory and Applications*. McGraw-Hill, New York, 1985.
12. Müller, P.C. and Schiehlen, W.: *Linear Vibrations*. Martinus Nijhoff, Dordrecht, 1985, Reprint Springer, Berlin.
13. Kurz, T., Eberhard, P., Henninger, C., and Schiehlen, W.: From Neweul to Neweul-M²: Symbolical Equations of Motion for Multibody System Analysis and Synthesis, Multibody System Dynamics, 24 (2010), 25–41.
14. Hollerbach, J.M.: A recursive Lagrangian formulation of manipulator dynamics and comparative study of dynamics formulation complexity. *IEEE Transactions on Systems, Man, Cybernetics*, 11 (1980), 730–736.
15. Bae, D.S. and Haug, E.J.: A recursive formulation for constrained mechanical system dynamics: Part I, open loop systems, *Mechanics of Structures and Machines*, 15 (1987), 359–382.

16. Brandl, H., Johanni, R., and Otter, M.: A very efficient algorithm for the simulation of robots and similar multibody systems without inversion of the mass matrix. In: P. Kopacek, I. Troch, and K. Desoyer (eds.). *Theory of Robots*. Pergamon, Oxford, 1988, pp. 95–100.
17. Schiehlen, W.: Computational aspects in multibody system dynamics. *Computer Methods in Applied Mechanics and Engineering*, 90 (1991), 569–582.
18. Bae, D.S. and Haug, E.J.: A recursive formulation for constrained mechanical system dynamics: Part II, closed loop systems. *Mechanics of Structures and Machines*, 15 (1987), 481–506.
19. Saha, S.K. and Schiehlen, W.: Recursive kinematics and dynamics for parallel structural closed-loop multibody systems. *Mechanics of Structures and Machines*, 29 (2001), 143–175.

3 Simulation Algorithms and Software Tools

Martin Arnold

CONTENTS

3.1 Introduction 45
3.2 Basic Simulation Algorithms 46
3.2.1 Static Analysis 47
3.2.2 Linearization 48
3.2.3 Time Integration 48
3.3 General Structure of Model Equations 50
3.3.1 Flexible Multi-Body Systems 50
3.3.2 Inner State Variables 51
3.3.3 Time Events 51
3.3.4 Kinematically Closed Loops 52
3.3.5 Advanced DAE Models 53
3.3.6 Summary 55
3.4 DAE Time Integration in Vehicle System Dynamics 55
3.4.1 DAE Time Integration by ODE Methods 56
3.4.2 Index Reduction and Projection 58
3.4.3 Gear–Gupta–Leimkuhler Formulation 60
3.4.4 Advanced DAE Models 60
3.4.5 Summary 61
3.5 Algorithms and Tools for the Simulation of Multi-Physical Problems 62
3.5.1 Extensions of Multi-Body System Simulation Packages 62
3.5.2 General Purpose Simulation Tools and Modeling Languages 64
3.5.3 Simulator Coupling 64
3.5.4 Summary 66
Acknowledgments 66
References 66

3.1 INTRODUCTION

Model-based simulation in vehicle system dynamics relies on advanced methods for model setup, robust and efficient numerical solution techniques, and powerful simulation tools for industrial applications.

Most frequently, the mechanical components are described by rigid or flexible multi-body system models that interact with electrical, hydraulic, and other system components [1,2] (see also [3] for a summary of recent developments and [4] for a comprehensive overview on methods and algorithms that are tailored to industrial simulation packages). The first systematic treatment of numerical

problems and numerical methods in multi-body dynamics from the viewpoint of vehicle system dynamics is the monograph of Eich-Soellner and Führer [5] (see also [6]).

In its simplest form, the equations of motion of a multi-body system are given by a nonlinear second-order system of ordinary differential equations (ODEs) of moderate dimension that may be solved numerically by standard methods.

The structure of the model equations gets more complicated if the multi-body system has kinematically *closed loops* [7]. Multi-body system models of wheel suspensions may be considered as a typical example in road vehicle simulation. The complex kinematics is modeled by *constraints*; the resulting model equations form a second-order differential-algebraic equation (DAE).

Robust and efficient DAE time integration methods are the backbone of state-of-the-art multi-body system simulation software. Another important issue in simulation is the handling of *discrete* controllers that are beyond the classical time continuous world of mechanics.

Advanced multi-body system simulation packages have been open to multidisciplinary applications for a long time. The classical approach to the integration of hydraulic and electrical system components in multi-body models are special force elements with time continuous or time discrete inner state variables [3].

Alternatively, novel modeling and simulation techniques for multi-physical technical systems may be used in more complex multidisciplinary applications. Here, the multidisciplinary aspect is considered by universal modeling languages like Modelica [8] or by the coupling of several monodisciplinary simulation packages in a co-simulation environment [9].

In this chapter, we give an overview on classical and more advanced simulation algorithms in vehicle system dynamics that are closely related to multi-body numerics and multi-body system simulation packages.

The material is organized as follows: Basic simulation algorithms in multi-body system analysis are summarized in Section 3.2. In Section 3.3, we discuss the general structure of model equations including closed-loop systems, inner state variables of force elements, and time discrete subsystems. Special DAE time integration methods for the robust and efficient numerical solution of these model equations are presented in Section 3.4. Finally, in Section 3.5, we focus on algorithms and tools for multidisciplinary applications.

Throughout this chapter, distributed physical phenomena like the elastic deformation of car components or the temperature field in disk brakes are considered only by low-dimensional modal approximations. Modeling and numerical solution by finite element techniques are studied separately in the subsequent chapter on FEM (see Chapter 4).

3.2 BASIC SIMULATION ALGORITHMS

If a mechanical multi-body system is described by a minimum set of generalized coordinates $\boldsymbol{y}(t) \in \mathbb{R}^{n_y}$, the equations of motion form a second-order ODE

$$\boldsymbol{M}(\boldsymbol{y})\ddot{\boldsymbol{y}}(t) = \boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) \tag{3.1}$$

with the symmetric, positive definite mass matrix $\boldsymbol{M}(\boldsymbol{y})$ containing mass and inertia properties of all bodies and the vector of applied and gyroscopic forces $\boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t)$ [7,10].

Despite its implicit structure, Equation (3.1) can be solved numerically with a complexity that grows only linearly with the number N of bodies in the system. These $\mathcal{O}(N)$ *formalisms* exploit the topology of the multi-body system to evaluate the right-hand side $\boldsymbol{M}^{-1}(\boldsymbol{y})\boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t)$ for given $t, \boldsymbol{y}, \dot{\boldsymbol{y}}$ (*explicit* formalisms, for example, [11,12]) or the residual $\boldsymbol{r}(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{a}, t) := \boldsymbol{M}(\boldsymbol{y})\boldsymbol{a} - \boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t)$ for given $t, \boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{a}$ (*residual* formalisms, for example, [13]) efficiently.

In contrast to structural dynamics, the standard time integration methods of multi-body dynamics are not tailored to the second-order structure of (3.1) since additional first-order equations

have to be considered frequently in practical applications (see Section 3.3). With the velocity vector $\boldsymbol{v} := \dot{\boldsymbol{y}}$, the equations of motion get the state space form

$$\dot{\boldsymbol{x}}(t) = \varphi(\boldsymbol{x}, t) \quad \text{with } \boldsymbol{x}(t) := \begin{pmatrix} \boldsymbol{y}(t) \\ \boldsymbol{v}(t) \end{pmatrix}, \quad \varphi(\boldsymbol{x}, t) := \begin{pmatrix} \boldsymbol{v} \\ [\boldsymbol{M}^{-1}\boldsymbol{f}](\boldsymbol{y}, \boldsymbol{v}, t) \end{pmatrix}. \tag{3.2}$$

3.2.1 Static Analysis

The computation of static equilibria is often the first step in the analysis of a multi-body system model. They define, for example, the working point for the linearization of (3.1) and provide initial values for the dynamical simulation. An equilibrium position $\boldsymbol{y}^*$ is determined by the equilibrium conditions $\dot{\boldsymbol{y}} = \ddot{\boldsymbol{y}} = \boldsymbol{0}$ at $t = t_0$:

$$\boldsymbol{0} = \boldsymbol{f}(\boldsymbol{y}^*, \boldsymbol{0}, t_0). \tag{3.3}$$

Equation (3.3) form a system of n_y nonlinear equations for the n_y unknowns $\boldsymbol{y}^*$ that is solved by Newton's method [14] being available as free software in a number of implementations that meet the demands on robustness and efficiency in engineering applications, see, for example, the results of the MINPACK project [15] at http://www.netlib.org/minpack.

Alternatively, the equilibrium position $\boldsymbol{y}^*$ of a stable system may also be obtained as limit $\boldsymbol{y}^* = \lim_{t\to\infty}\boldsymbol{y}(t)$ by a long-term time integration of the equations of motion (3.1) without external excitations. The time integration to the equilibrium position needs more computing time than Newton's method but in critical applications it is often the more robust approach.

Problems with the same mathematical structure as the equilibrium equations (3.3) result from the analysis of steady states. If a car runs with constant speed v_0 in a constant radius path, its position and velocity along this path are given by $s(t) = s_0 + v_0(t - t_0)$ and by v_0, respectively, and a force acts in normal direction to keep the radius constant. The remaining $n_y - 1$ components $\bar{\boldsymbol{f}}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t)$ of the force vector $\boldsymbol{f}$ in (3.1) have to vanish and give the first $n_y - 1$ equations of the n_y steady-state conditions

$$\boldsymbol{0} = \bar{\boldsymbol{f}}(\boldsymbol{y}^*, \boldsymbol{v}^*, t_0), \quad s_0 = s(t_0) \tag{3.4}$$

that have to be solved w.r.t. $\boldsymbol{y}^* \in \mathbb{R}^{n_y}$. All components of the steady-state velocity vector $\boldsymbol{v}^*$ are equal to zero except the one that is tangential to the constant radius path ($v_0^* = v_0$). For steady-state problems (3.4), Newton's method (or, alternatively, the long-term time integration of a stable system into its steady state) is again the method of choice.

There are some other static analysis methods that are closely related to the equilibrium problem (3.3): the computation of nominal forces in vehicle dynamics [5] and the trimming of aircraft models in aeronautics [16]. In both applications, the actual values of some parameters $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ in the force elements are left undefined in the model setup: $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t; \boldsymbol{\theta})$. Later, these parameters are adjusted to meet equilibrium conditions

$$\boldsymbol{0} = \boldsymbol{f}(\boldsymbol{y}_0, \boldsymbol{0}, t_0; \boldsymbol{\theta}^*) \tag{3.5}$$

for a predefined position $\boldsymbol{y}_0$ of the multi-body system. The actual values $\boldsymbol{\theta}^*$ of nominal forces or trim parameters are important criteria for the evaluation of a vehicle or aircraft design.

Equation (3.5) has the same mathematical structure as the classical equilibrium problem (3.3) and could in principle be solved again by Newton's method. However, in practical

applications, the system (3.5) is often underdetermined. Following the approach of Levenberg and Marquardt, (3.5) may be substituted by

$$\| \boldsymbol{f}(\boldsymbol{y}_0, \boldsymbol{0}, t_0; \boldsymbol{\theta}^*) \|_2 + \alpha \, \|\boldsymbol{\theta}^*\|_2 \rightarrow \min$$

to guarantee uniqueness of $\boldsymbol{\theta}^*$. Here, $\alpha > 0$ denotes a small scalar regularization parameter [17].

3.2.2 Linearization

The linearization of the equations of motion (3.1) is the key to the linear stability analysis near an equilibrium (see Chapter 5) and to other methods of linear system analysis [2]. Writing the right-hand side of (3.1) as $\boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{u}(t))$ with system inputs $\boldsymbol{u}(t) \in \mathbb{R}^{n_u}$, the linearization at an equilibrium $\boldsymbol{y} = \boldsymbol{y}^*, \dot{\boldsymbol{y}} = \boldsymbol{0}, \boldsymbol{u} = \boldsymbol{0}$ yields

$$\boldsymbol{M}\ddot{\bar{\boldsymbol{y}}} = -\boldsymbol{D}\dot{\bar{\boldsymbol{y}}} - \boldsymbol{K}\bar{\boldsymbol{y}} + \boldsymbol{B}\boldsymbol{u}(t) \tag{3.6}$$

with

$$\boldsymbol{M} = \boldsymbol{M}(\boldsymbol{y}^*), \quad \boldsymbol{D} = -\frac{\partial \boldsymbol{f}}{\partial \dot{\boldsymbol{y}}}(\boldsymbol{y}^*, \boldsymbol{0}, \boldsymbol{0}), \quad \boldsymbol{K} = -\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}(\boldsymbol{y}^*, \boldsymbol{0}, \boldsymbol{0}), \quad \boldsymbol{B} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(\boldsymbol{y}^*, \boldsymbol{0}, \boldsymbol{0}).$$

The Jacobians $\partial \boldsymbol{f}/\partial \boldsymbol{y}$, $\partial \boldsymbol{f}/\partial \dot{\boldsymbol{y}}$, and $\partial \boldsymbol{f}/\partial \boldsymbol{u}$ are approximated by classical finite differences [14]. We obtain for the ith column of $\partial \boldsymbol{f}/\partial \boldsymbol{y}$

$$\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}(\boldsymbol{y}^*, \boldsymbol{0}, \boldsymbol{0}) \right)_{\bullet, i} \approx \frac{\boldsymbol{f}(\boldsymbol{y}^* + \Delta \cdot \boldsymbol{e}_i, \boldsymbol{0}, \boldsymbol{0}) - \boldsymbol{f}(\boldsymbol{y}^*, \boldsymbol{0}, \boldsymbol{0})}{\Delta} \tag{3.7}$$

with the ith unit vector $\boldsymbol{e}_i$ and a small scalar parameter Δ that satisfies $0 < |\Delta| \ll 1$.

The numerical effort for the linearization (3.6) is dominated by the difference approximations (3.7) with a total of $2n_y + n_u + 1$ evaluations of $\boldsymbol{f}$ that require $2n_y + n_u + 1$ evaluations of each force element in the multi-body system model. Substantial savings are achieved by the simultaneous difference approximation of several columns of $\partial \boldsymbol{f}/\partial \boldsymbol{y}$, $\partial \boldsymbol{f}/\partial \dot{\boldsymbol{y}}$, and $\partial \boldsymbol{f}/\partial \boldsymbol{u}$ that requires some graph theoretical preparations based on the topology of the multi-body system [18,19].

3.2.3 Time Integration

The state space form (3.2) of the equations of motion may be integrated by any standard ODE solver [20]. Explicit Runge–Kutta methods like the fifth-order method of Dormand and Prince (free FORTRAN code **DOPRI5** [21] and default ODE solver **ode45** in MATLAB®,* see [22]) and predictor–corrector methods of Adam's type (free FORTRAN code **LSODE** at http://www.netlib.org/odepack and MATLAB solver **ode113**) proved to be favorable in non-stiff applications.

More frequently, the model equations in vehicle system dynamics appear to be *stiff* [20,23] because of stiff springs and/or strongly damping force elements in most multi-body vehicle models (for additional information on both non-stiff and stiff problems, see Chapter 5). Backward differentiation formulae (BDF, also Gear's method) are the most often used time

* MATLAB is a trademark of The MathWorks, Inc., http://www.mathworks.com.

integration methods for stiff technical systems. Starting from initial values $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$, the numerical solution is obtained time step by time step solving

$$\frac{1}{h_n}\sum_{j=0}^{k_n}\alpha_{n,j}\boldsymbol{x}_{n+1-j} = \boldsymbol{\varphi}(\boldsymbol{x}_{n+1}, t_{n+1}) \tag{3.8}$$

w.r.t. $\boldsymbol{x}_{n+1} \approx \boldsymbol{x}(t_{n+1})$, $(n = 0, 1, \ldots)$.

In (3.8), the stepsize of time step $t_n \to t_{n+1}$ is denoted by $h_n := t_{n+1} - t_n$. The parameter k_n defines the order of the method, $1 \le k_n \le 6$, and $\alpha_{n,j}$ are the BDF coefficients that are determined by k_n and by the stepsizes $h_n, h_{n-1}, \ldots$ [20]. The well-known backward Euler method

$$\frac{1}{h_n}(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) = \boldsymbol{\varphi}(\boldsymbol{x}_{n+1}, t_{n+1})$$

is a special case of (3.8), $k_n = 1$.

For given state vectors $\boldsymbol{x}_{n+1-j} \approx \boldsymbol{x}(t_{n+1-j})$, $(j = 1, \ldots, k)$, the BDF equations (3.8) define $\boldsymbol{x}_{n+1}$ implicitly as solution of the system of n_x nonlinear equations

$$\boldsymbol{0} = \Phi(\boldsymbol{x}) := \frac{\alpha_{n,0}}{h_n}\boldsymbol{x} - \boldsymbol{\varphi}(\boldsymbol{x}, t_{n+1}) + \frac{1}{h_n}\sum_{j=1}^{k_n}\alpha_{n,j}\boldsymbol{x}_{n+1-j}. \tag{3.9}$$

The application of Newton's method to (3.9) yields

$$\boldsymbol{x}_{n+1}^{(l+1)} := \boldsymbol{x}_{n+1}^{(l)} - \boldsymbol{J}^{-1}\cdot\boldsymbol{\Phi}(\boldsymbol{x}_{n+1}^{(l)}), \quad (l \ge 0) \tag{3.10}$$

with an initial guess $\boldsymbol{x}_{n+1}^{(0)}$ that is obtained by polynomial extrapolation of $\boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \ldots$. In (3.10), matrix $\boldsymbol{J}$ approximates the Jacobian of (3.9):

$$\boldsymbol{J} \approx \frac{\partial\boldsymbol{\Phi}}{\partial\boldsymbol{x}}\left(\boldsymbol{x}_{n+1}^{(l)}\right) = \frac{\alpha_{n,0}}{h_n}\boldsymbol{I}_{n_x} - \frac{\partial\boldsymbol{\varphi}}{\partial\boldsymbol{x}}\left(\boldsymbol{x}_{n+1}^{(l)}, t_{n+1}\right). \tag{3.11}$$

Typically, the approximation of $\partial\boldsymbol{\varphi}/\partial\boldsymbol{x}$ by finite differences according to (3.7) dominates the overall computing time in the dynamical simulation because of the large number of function evaluations of $\boldsymbol{\varphi}$ that require the evaluation of all force elements in the multi-body system and the computation of $[\boldsymbol{M}^{-1}\boldsymbol{f}](\boldsymbol{y}, \boldsymbol{v}, t)$ by the O(N) formalism (see 3.2). Therefore, these time-consuming reevaluations of the Jacobian $\boldsymbol{J}$ have to be avoided as far as possible [19,20].

In practical computations, stepsize h_n and order k_n are adapted automatically to meet the user-defined error tolerances. The combination of stepsize and order control with the algorithms for Newton's method (3.10) and for the Jacobian reevaluations results in a fairly complicated structure of state-of-the-art BDF solvers. Free FORTRAN codes are **LSODE** at http://www.netlib.org/odepack and **DASSL** at http://www.netlib.org/ode. The MATLAB BDF solver is **ode15s** (see [22]).

From the viewpoint of multi-body dynamics, the solver **DASSL** [24] is especially interesting because of its interface in residual form

$$\boldsymbol{0} = \boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) \tag{3.12}$$

that is a generalization of (3.2) with $\boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) = \dot{\boldsymbol{x}} - \boldsymbol{\varphi}(\boldsymbol{x}, t)$ and optionally allows also the use of very efficient residual formalisms [13]:

$$\boldsymbol{x}(t) := \begin{pmatrix}\boldsymbol{y}(t)\\ \boldsymbol{v}(t)\end{pmatrix}, \quad \boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) := \begin{pmatrix}\dot{\boldsymbol{y}} - \boldsymbol{v}\\ \boldsymbol{M}(\boldsymbol{y})\dot{\boldsymbol{v}} - \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{v}, t)\end{pmatrix}.$$

Furthermore, **DASSL** may also be used in the time integration of closed-loop systems (see Section 3.4).

BDF for (3.12) have the form

$$\mathbf{0} = \boldsymbol{F}\left(\boldsymbol{x}_{n+1},\ \frac{1}{h_n}\sum_{j=0}^{k_n} \alpha_{n,j}\boldsymbol{x}_{n+1-j},\ t_{n+1} \right) \tag{3.13}$$

with the Jacobian

$$\boldsymbol{J} \approx \frac{\alpha_{n,0}}{h_n}\frac{\partial \boldsymbol{F}}{\partial \dot{\boldsymbol{x}}}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) \tag{3.14}$$

instead of (3.11). To keep the efficiency of classical BDF solvers in the application to multi-body systems, **DASSL**'s standard algorithm for Jacobian reevaluation has to be modified to exploit the special block structure of $\partial \boldsymbol{F}/\partial \dot{\boldsymbol{x}}$ (see [19]):

Explicit formalism: $\dfrac{\partial \boldsymbol{F}}{\partial \dot{\boldsymbol{x}}} = \begin{pmatrix} \boldsymbol{I}_{n_y} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n_y} \end{pmatrix}$, residual formalism: $\dfrac{\partial \boldsymbol{F}}{\partial \dot{\boldsymbol{x}}} = \begin{pmatrix} \boldsymbol{I}_{n_y} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M} \end{pmatrix}$.

3.3 GENERAL STRUCTURE OF MODEL EQUATIONS

The numerical methods that are discussed in Section 3.2 may be extended from the equations of motion (3.1) to more complex model equations that are typical of applications in vehicle system dynamics.

3.3.1 Flexible Multi-Body Systems

The elastic deformation of flexible bodies is not covered by the classical equations of motion (3.1) of rigid multi-body systems. In the moving frame of reference formulation [25–27], the configuration $\vec{u}^{(i)}(\vec{x}, t)$ of a flexible body $(.)^{(i)}$ in $\mathbb{R}^3$ is represented by the gross motion of a body fixed reference frame with coordinates $\boldsymbol{y}_r^{(i)}(t)$ and the (small) elastic deformation $\vec{w}^{(i)}(\vec{x}, t)$ w.r.t. this reference configuration:

$$\vec{u}^{(i)}(\vec{x}, t) = \vec{r}\left(\boldsymbol{y}_r^{(i)}(t) \right) + \vec{w}^{(i)}(\vec{x}, t). \tag{3.15}$$

Following a Rayleigh–Ritz approach, a low-dimensional modal approximation of $\vec{w}^{(i)}(\vec{x}, t)$ is used:

$$\vec{w}^{(i)}(\vec{x}, t) \approx \sum_{j=1}^{\sigma_i} q_j^{(i)}(t)\, \vec{w}_j^{(i)}(\vec{x}). \tag{3.16}$$

In general, the σ_i modes $\vec{w}_j^{(i)}$ are selected based on a finite element analysis of the flexible body $(.)^{(i)}$. Typical mode functions $\vec{w}_j^{(i)}$ are eigenmodes in the frequency range up to 50–100 Hz, static modes, or frequency response modes [27,28].

With (3.15) and (3.16), the equations of motion for flexible multi-body systems get the basic structure (3.1) with a coordinate vector $\boldsymbol{y}(t)$ that contains the generalized coordinates of rigid bodies and the elastic coordinates $\boldsymbol{y}_r^{(i)}(t)$ and $\left(q_1^{(i)}(t), \ldots, q_{\sigma_i}^{(i)}(t) \right)$ of flexible bodies. $\boldsymbol{M}(\boldsymbol{y})$ and $\boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t)$ are extended by the modal mass, damping and stiffness matrices and by the coupling terms between gross motion and elastic deformation $\vec{w}^{(i)}$ (see [26,29]). In industrial applications, these data are generated automatically using standardized interfaces to finite element tools [30].

Today, the selection of an appropriate set of modes $\left\{ \vec{w}_1^{(i)}, \ldots, \vec{w}_{\sigma_i}^{(i)} \right\}$ still relies much on engineering intuition. Algorithms for (semi)automatic mode selection are a topic of active research (see [28]

for a criterion that is based on a linear error analysis, [29] for a more general approach that could be used in an adaptive mode selection strategy, and [31] for the use of *thermal response* modes in thermoelastic applications).

Recently, the moving frame of reference formulation has been extended to a modal multifield approach that considers elastic deformation as well as electrostatic fields in piezo elements and spatial temperature distributions in thermoelastic problems like the thermal deformation of disk brakes [31].

3.3.2 Inner State Variables

Frequently, force elements in multi-body system models describe engineering systems that have their own, internal dynamics (*dynamical* force elements [5]). Typical examples include hydraulic components, control devices (see Chapters 7 and 22), and advanced tire models (see Chapters 17 and 18). From the mathematical viewpoint, these force elements are characterized by *inner* state variables. Elements like hydraulic components have time continuous state variables $\boldsymbol{c}(t)$ with state equations

$$\dot{\boldsymbol{c}}(t) = \boldsymbol{d}(\boldsymbol{c}, \boldsymbol{y}, \dot{\boldsymbol{y}}, t). \tag{3.17}$$

The internal dynamics of discrete controllers is characterized by time discrete state variables $\boldsymbol{r}$ with $\boldsymbol{r} = \boldsymbol{r}_j$ in the sampling interval $[T_j, T_{j+1})$ and

$$\boldsymbol{r}_{j+1} = \boldsymbol{k}(\boldsymbol{r}_j, \boldsymbol{r}_{j-1}, \ldots, \boldsymbol{y}, \dot{\boldsymbol{y}}, T_{j+1}). \tag{3.18}$$

In most applications, the sampling points are equally spaced: $T_j = j \cdot \Delta_t$, typical sampling rates are in the range of Δ_t = 1–10 ms.

In the multi-body system, these force elements contribute to the force vector $\boldsymbol{f}$ in (3.1):

$$\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{y}(t), \dot{\boldsymbol{y}}(t), \boldsymbol{c}(t), \boldsymbol{r}_j, t) \quad \text{if } t \in [T_j, T_{j+1}). \tag{3.19}$$

With (3.18) and (3.19), the equations of motion are not longer continuous in time but form a hybrid system with continuous and discrete components. In the dynamical simulation, the variable order, variable stepsize integrators of Section 3.2 may be used only within a single sampling interval $[T_j, T_{j+1}]$. At $t = T_{j+1}$, the integration has to be stopped, the discrete variables $\boldsymbol{r}$ are updated according to (3.18), and the time integration may be continued for $t \in [T_{j+1}, T_{j+2}]$.

The frequent re-initializations of the ODE solvers have a very negative effect on their performance that is compensated in part by an adapted algorithm for Jacobian reevaluations [4,19]. Alternatively, special Runge–Kutta type methods are proposed for a more efficient re-initialization after the sampling point T_{j+1} (see [32]).

3.3.3 Time Events

The discontinuous changes (3.18) of state variables are special *time events* [20]. The term "time event" refers to a discontinuity ("event") in state variables or model equations at an isolated time $t = T^*$ during simulation. For model equations (3.1) with $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{c}, \boldsymbol{r}_j, \boldsymbol{u}(t); \boldsymbol{\theta})$, typical time events result from switching processes with jumping system inputs $\boldsymbol{u}(t)$ and from contact problems including stick-slip phenomena with force elements that change their force parameters $\boldsymbol{\theta}$ or even their internal structure during simulation (unilateral force elements, varying friction coefficients $\boldsymbol{\theta}$, etc. [5,33], see also Chapter 5).

For switching inputs $\boldsymbol{u}(t)$ the time T^* is known in advance but, in general, T^* has to be determined during simulation. State-of-the-art ODE solvers locate time events automatically using *switching functions* that have to be provided in addition to the model equations (3.1) (see [5,24]).

In case of a time event, the time integration is stopped at $t = T^*$, system variables $\boldsymbol{y}$, $\boldsymbol{c}$, $\boldsymbol{r}_j$ and system parameters $\boldsymbol{\theta}$ are updated (if necessary), and the simulation is continued after a re-initialization of the solver. Again, frequent time events may slow down the time integration substantially.

3.3.4 Kinematically Closed Loops

In the setup of multi-body system models, it is important to select an appropriate set of coordinates $\boldsymbol{y}$. For tree-structured systems, joint coordinates $\boldsymbol{y}$ that describe the degrees of freedom in the joints of the multi-body system form a minimum set of generalized coordinates resulting in equations of motion in ODE form (3.1).

If, however, the multi-body system has kinematically closed loops, then the joint coordinates $\boldsymbol{y}$ are no longer independent of each other. Loop closing joints restrict the system's configuration to joint coordinates $\boldsymbol{y}$ satisfying n_g *constraints*

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, t). \tag{3.20}$$

Coordinate partitioning methods [7,34] locally select a subset of linearly independent joint coordinates and compute the remaining (dependent) joint coordinates solving the system of nonlinear equations (3.20). From the numerical viewpoint, it proved to be more favorable to keep *all* joint coordinates in the equations of motion that get the form

$$\boldsymbol{M}(\boldsymbol{y})\,\ddot{\boldsymbol{y}}(t) = \boldsymbol{f}(\boldsymbol{y},\,\dot{\boldsymbol{y}},\,t) - \boldsymbol{G}^{\top}(\boldsymbol{y},\,t)\lambda, \tag{3.21a}$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, t) \tag{3.21b}$$

with constraint forces $-\boldsymbol{G}^{\top}(\boldsymbol{y},\,t)\boldsymbol{\lambda}$ that are determined by the constraint matrix $\boldsymbol{G}(\boldsymbol{y},\,t)$: = $(\partial \boldsymbol{g}/\partial \boldsymbol{y})$ $(\boldsymbol{y},\,t)$ and the Lagrange multipliers $\boldsymbol{\lambda}(t) \in \mathbb{R}^{n_g}$ (see [12]). The common assumption rank $\boldsymbol{G}(\boldsymbol{y},\,t) = n_g$ (*Grübler* condition) excludes redundant constraints (3.21b). Equations (3.21) form a second-order DAE, the *descriptor form* of the equations of motion [24].

The descriptor form is not restricted to joint coordinates $\boldsymbol{y}$. Using absolute (Cartesian) coordinates $\boldsymbol{y}$, *any* joint results in a constraint (3.21b). The resulting equations of motion (3.21) have a much larger dimension than in the joint coordinate case but $\boldsymbol{M}(\boldsymbol{y})$, $\partial \boldsymbol{f}/\partial \boldsymbol{y}$, $\partial \boldsymbol{f}/\partial \dot{\boldsymbol{y}}$, $\boldsymbol{G}^{\top}$ are sparse matrices with a structure that reflects the topology of the multi-body system [35]. The counterpart to O(N) formalisms that eliminate redundant coordinates in a joint coordinate formulation [11,12] are *topological solvers* for the absolute coordinate formulation that allow an efficient elimination of redundant coordinates in (3.10) (see [32,35]).

In DAE terminology, the descriptor form (3.21) is an index-3 system (see [23,24]). The constraints (3.21b) imply additional restrictions on the state variables $\boldsymbol{y}(t)$ and $\boldsymbol{\lambda}(t)$. These *hidden* constraints are obtained by differentiation of (3.21b):

$$\begin{aligned}\boldsymbol{0} &= \frac{\mathrm{d}}{\mathrm{d}t}\,\boldsymbol{g}(\boldsymbol{y}(t), t) = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}}(\boldsymbol{y}, t)\cdot\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}(t) + \frac{\partial \boldsymbol{g}}{\partial t}(\boldsymbol{y}, t)\\ &= \boldsymbol{G}(\boldsymbol{y}, t)\,\dot{\boldsymbol{y}}(t) + \boldsymbol{g}^{(I)}(\boldsymbol{y}, t),\end{aligned} \tag{3.22a}$$

$$\boldsymbol{0} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\boldsymbol{g}(\boldsymbol{y}(t), t) = \boldsymbol{G}(\boldsymbol{y}, t)\,\ddot{\boldsymbol{y}}(t) + \boldsymbol{g}^{(II)}(\boldsymbol{y},\,\dot{\boldsymbol{y}},\,t) \tag{3.22b}$$

with functions $\boldsymbol{g}^{(I)}$ and $\boldsymbol{g}^{(II)}$ that summarize higher-order derivatives of $\boldsymbol{g}$ and the partial derivatives w.r.t. t. The solution of (3.21) has to satisfy the constraints (3.21b) on position level as well as the hidden constraints (3.22) on velocity and acceleration level.

Initial values $\boldsymbol{y}_0, \dot{\boldsymbol{y}}_0, \boldsymbol{\lambda}_0$ have to be *consistent* with all these constraints:

$$\begin{aligned} \mathbf{0} &= \boldsymbol{g}(\boldsymbol{y}_0, t_0), \quad \mathbf{0} = \boldsymbol{G}(\boldsymbol{y}_0, t_0)\,\dot{\boldsymbol{y}}_0 + \boldsymbol{g}^{(I)}(\boldsymbol{y}_0, t_0), \\ \mathbf{0} &= \boldsymbol{G}(\boldsymbol{y}_0, t_0)\,\ddot{\boldsymbol{y}}_0 + \boldsymbol{g}^{(II)}(\boldsymbol{y}_0, \dot{\boldsymbol{y}}_0, t_0) \end{aligned} \tag{3.23}$$

with $\boldsymbol{M}(\boldsymbol{y}_0)\,\ddot{\boldsymbol{y}}_0(t) = \boldsymbol{f}(\boldsymbol{y}_0, \dot{\boldsymbol{y}}_0, t_0) - \boldsymbol{G}^{\top}(\boldsymbol{y}_0, t_0)\boldsymbol{\lambda}_0$.

The computation of consistent initial values $\boldsymbol{y}_0, \dot{\boldsymbol{y}}_0, \boldsymbol{\lambda}_0$ is part of the initialization in the time integration of (3.21). If initial values $y_{0,ij}$ are given for $n_y - n_g$ independent position coordinates y_{ij}, $(j = 1, \ldots, n_y - n_g)$, the remaining initial values $y_{0,ij}$, $(j = n_y - n_g + 1, \ldots, n_y)$ are uniquely determined by (3.23) and may be computed by Newton's method. Furthermore, (3.23) defines systems of linear equations for the initial values $\dot{\boldsymbol{y}}_{0,ij}$, $(j = n_y - n_g + 1, \ldots, n_y)$ of n_g-dependent velocity components and for the initial values $\boldsymbol{\lambda}_0$.

3.3.5 Advanced DAE Models

The descriptor form (3.21) is linear w.r.t. $\boldsymbol{\lambda}$. This structural property is lost if the model has friction forces that depend on the constraint forces $-\boldsymbol{G}^{\top}\boldsymbol{\lambda}$. The resulting force vector $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{\lambda}, t)$ has to satisfy the generalized Grübler condition

$$\operatorname{rank} \boldsymbol{G}(\boldsymbol{y}, t) = \operatorname{rank}\left(\boldsymbol{G}(\boldsymbol{y}, t)\,\boldsymbol{M}^{-1}(\boldsymbol{y})\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\lambda}}(\boldsymbol{y}, \dot{\boldsymbol{y}}, \boldsymbol{\lambda}, t) - \boldsymbol{G}^{\top}(\boldsymbol{y}, t) \right) \right) = n_g \tag{3.24}$$

to guarantee the unique solvability of (3.21).

Another extension of (3.21) shows the large potential of DAE formulations in view of user-friendly algorithms for model setup. Consider rigid bodies $(.)^{(i)}$ and $(.)^{(j)}$ that are in contact. To formulate the contact conditions, parameterizations $\boldsymbol{s}^{(i)}, \boldsymbol{s}^{(j)} \in \mathbb{R}^2$ of the body surfaces are introduced. Then the position of the points $\vec{P}^{(i)}$ on the surface of body $(.)^{(i)}$ in $\mathbb{R}^3$ is determined by the position and orientation of this body, that is, by the multi-body system coordinates $\boldsymbol{y}$, and by the surface parameter $\boldsymbol{s}^{(i)}$: $\vec{P}^{(i)} = \vec{P}^{(i)}(\boldsymbol{y}, \boldsymbol{s}^{(i)})$ (see Figure 3.1).

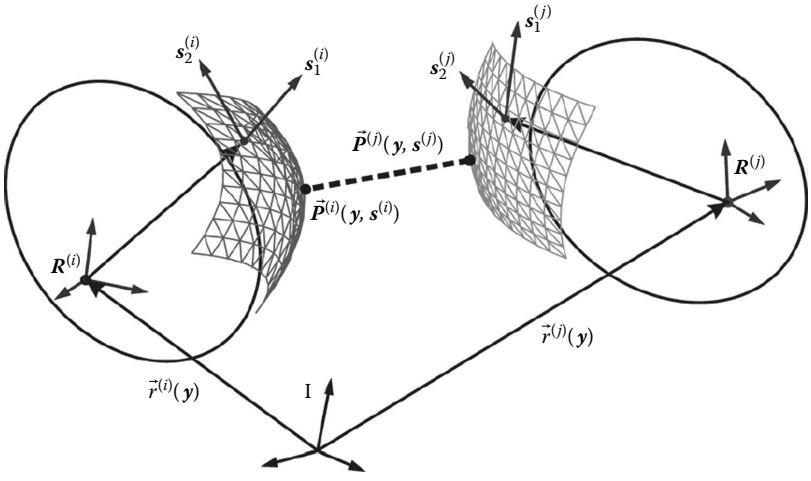


FIGURE 3.1 Formulation of contact conditions using surface parameters $\boldsymbol{s}$ as additional algebraic variables.

If both bodies are strictly convex, then there is a uniquely defined contact point $\vec{P}^{(i)}(\boldsymbol{y}, \boldsymbol{s}^{(i)}) = \vec{P}^{(j)}(\boldsymbol{y}, \boldsymbol{s}^{(j)})$ that belongs to the surfaces of both bodies. The contact point is characterized by the condition that the normal vector of body $(.)^{(i)}$ at $\vec{P}^{(i)}$ is parallel to the normal vector of body $(.)^{(j)}$ at $\vec{P}^{(j)}$ (see [12]). This geometrical condition defines implicitly the contact point coordinates $\boldsymbol{s} = (\boldsymbol{s}^{(i)}, \boldsymbol{s}^{(j)})$ as solution $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{y}, t)$ of a nonlinear system

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}, t) \tag{3.25}$$

with non-singular Jacobian $\partial \boldsymbol{h}/\partial \boldsymbol{s}$ (see [5,6]).

Contact and friction forces act at the contact point. Furthermore, the contact point coordinates $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{y}, t)$ are used in the formulation of the contact condition

$$\boldsymbol{0} = \vec{P}^{(i)}(\boldsymbol{y}, \boldsymbol{s}^{(i)}) - \vec{P}^{(j)}(\boldsymbol{y}, \boldsymbol{s}^{(j)}) \tag{3.26}$$

that defines constraints $\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, t)$ in (3.21). In a classical implementation, each evaluation of $\boldsymbol{g}(\boldsymbol{y}, t)$ would require the solution of (3.25) to get the contact point coordinates $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{y}, t)$ that have to be inserted in the contact condition (3.26).

As an alternative, the DAE framework allows to add simply the nonlinear equations (3.25) to the equations of motion (3.21) and to append the contact conditions (3.26) directly to the constraints:

$$\boldsymbol{M}(\boldsymbol{y})\,\ddot{\boldsymbol{y}}(t) = \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \lambda, t) - \boldsymbol{G}^{\top}(\boldsymbol{y}, \boldsymbol{s}, t)\boldsymbol{\lambda}, \tag{3.27a}$$

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}, t), \tag{3.27b}$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{s}, t). \tag{3.27c}$$

With this formulation, the solution of (3.25) w.r.t. $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{y}, t)$ is left to the DAE solver [4,6]. Note that the constraint matrix is now

$$\begin{aligned} \boldsymbol{G}(\boldsymbol{y}, \boldsymbol{s}, t) &= \frac{\mathrm{D}}{\mathrm{D}\boldsymbol{y}}\,\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{s}(\boldsymbol{y}, t), t) = \left[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}} + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{s}}\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{y}}\right](\boldsymbol{y}, \boldsymbol{s}, t) \\ &= \left[\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}} - \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{s}}\left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{s}}\right)^{-1}\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{y}}\right](\boldsymbol{y}, \boldsymbol{s}, t) \end{aligned} \tag{3.28}$$

because of

$$\boldsymbol{0} = \frac{\mathrm{D}}{\mathrm{D}\boldsymbol{y}}\,\boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}(\boldsymbol{y}, t), t) = \left[\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{y}} + \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{s}}\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{y}}\right](\boldsymbol{y}, \boldsymbol{s}, t). \tag{3.29}$$

From the mathematical viewpoint, Equation (3.25) is just a special case of algebraic equations

$$\boldsymbol{0} = \boldsymbol{b}(\boldsymbol{w}, \boldsymbol{y}, \dot{\boldsymbol{y}}, \lambda, t) \tag{3.30}$$

with variables $\boldsymbol{w}(t) \in \mathbb{R}^{n_b}$ that are defined implicitly by (3.30) if $\partial \boldsymbol{b}/\partial \boldsymbol{w}$ is non-singular. The mathematical modeling of force elements may be simplified substantially using such auxiliary variables $\boldsymbol{w}$. A typical example is the modeling of joint friction (see [4] for a detailed discussion).

With variables $\boldsymbol{w}$ the force vector in (3.27) gets the general form $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{\lambda}, \boldsymbol{w}, t)$ and the Jacobian $\partial \boldsymbol{f}/\partial \boldsymbol{\lambda}$ in the generalized Grübler condition (3.24) has to be substituted by

$$\left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\lambda}} - \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}} \left(\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{w}} \right)^{-1} \frac{\partial \boldsymbol{b}}{\partial \lambda} \right] (\boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{\lambda}, \boldsymbol{w}, t).$$

3.3.6 Summary

Based on the multi-body system approach, the model equations in vehicle system dynamics are formulated as hybrid system of DAEs

$$\boldsymbol{M}(\boldsymbol{y})\, \ddot{\boldsymbol{y}}(t) = \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{c}, \boldsymbol{r}_j, \boldsymbol{\lambda}, \boldsymbol{w}, t) - \boldsymbol{G}^{\top}(\boldsymbol{y}, \boldsymbol{s}, t)\boldsymbol{\lambda}, \tag{3.31a}$$

$$\dot{\boldsymbol{c}}(t) = \boldsymbol{d}(\boldsymbol{c}, \boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{r}_j, \boldsymbol{\lambda}, \boldsymbol{w}, t). \tag{3.31b}$$

$$\boldsymbol{0} = \boldsymbol{b}(\boldsymbol{w}, \boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{c}, \boldsymbol{r}_j, \boldsymbol{\lambda}, t), \tag{3.31c}$$

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}_j, t), \tag{3.31d}$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{r}_j, t) \tag{3.31e}$$

with $t \in [T_j, T_{j+1})$ and discrete state equations

$$\boldsymbol{r}_{j+1} = \boldsymbol{k}(\boldsymbol{r}_j, \boldsymbol{r}_{j-1}, \ldots, \boldsymbol{y}, \boldsymbol{s}, \dot{\boldsymbol{y}}, \boldsymbol{c}, \boldsymbol{\lambda}, \boldsymbol{w}, T_{j+1}). \tag{3.32}$$

The continuous part (3.31) of the model equations forms an index-3 DAE with a complex but characteristic structure that will be exploited in the numerical solution (see Section 3.4).

It is important that the contact point coordinates $\boldsymbol{s}$ in the constraints (3.31e) are defined by an equation that is independent of $\dot{\boldsymbol{y}}$ and $\boldsymbol{\lambda}$ (see 3.31d). Otherwise, the model equations would not be solvable unless strong additional regularity conditions were satisfied.

3.4 DAE TIME INTEGRATION IN VEHICLE SYSTEM DYNAMICS

The time integration of DAE model equations (3.31) is based on classical ODE methods combined with *index reduction* and *projection* techniques. The time integration methods will be discussed in detail for DAEs (3.21) that are rewritten as first-order systems in residual form (3.12) with

$$\boldsymbol{x}(t) = \begin{pmatrix} \boldsymbol{y}(t) \\ \boldsymbol{v}(t) \\ \boldsymbol{\lambda}(t) \end{pmatrix}, \quad \boldsymbol{0} = \boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) = \begin{pmatrix} \dot{\boldsymbol{y}} - \boldsymbol{v} \\ \boldsymbol{M}(\boldsymbol{y})\dot{\boldsymbol{v}} - \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{v}, t) + \boldsymbol{G}^{\top}(\boldsymbol{y}, t)\boldsymbol{\lambda} \\ \boldsymbol{g}(\boldsymbol{y}, t) \end{pmatrix}. \tag{3.33}$$

At the end of this section, we will come back to DAEs of the more general form (3.31).

The first generation of DAE time integration methods failed in the application to constrained mechanical systems (see [36]). Very similar numerical problems were observed in quite different practical applications like robotics, road vehicle simulation, and railway vehicle simulation. To discuss

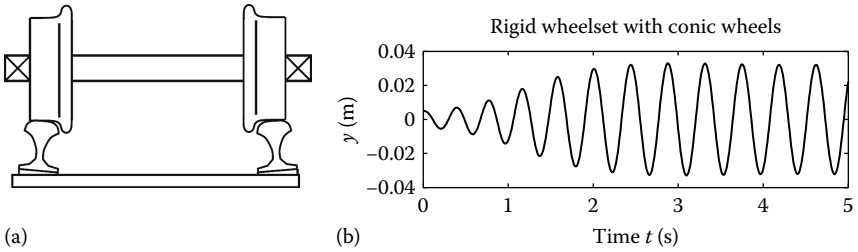


FIGURE 3.2 (a) Rigid wheelset performing a hunting motion, (b) Lateral displacement y.

these numerical effects in more detail, one of the best analyzed benchmark problems from the literature will be used throughout this section [6,37]. This multi-body system model comes from railway dynamics, but in time integration it shows all those numerical effects that are typical of any constrained systems (3.33) including the ones from road vehicle dynamics.

The benchmark describes a rigid wheelset with conic wheels moving with constant speed along a straight track. Starting with a small initial lateral displacement y, the wheelset oscillates in lateral direction (see Figure 3.2). This so-called hunting motion is a well-known phenomenon in railway vehicle dynamics.

The equations of motion are formulated in DAE form (3.27) with the position coordinates $\boldsymbol{y}(t) \in \mathbb{R}^6$ of the wheelset that correspond to its six degrees of freedom. Two constraints (3.27c) guarantee permanent contact between the two wheels and the rails [6]. Exploiting symmetries in the geometry of wheel and rail, the contact point between a single wheel and the rail may be described by a scalar position coordinate $s(t)$ resulting in two equations (3.27b) that determine implicitly $\boldsymbol{s} = (s_l, s_r)^\top \in \mathbb{R}^2$ for the contact point coordinates on left and right wheel [37].

3.4.1 DAE Time Integration by ODE Methods

Equation (3.33) suggests to apply BDF (3.13) and other ODE methods straightforwardly to the time integration of DAEs. Simple test problems like (3.33) with $\boldsymbol{M} = \boldsymbol{I}_{n_y}$, $\boldsymbol{f} = \boldsymbol{0}$, and linear constraints

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, t) = \boldsymbol{C}\boldsymbol{y} - \boldsymbol{z}(t)$$

show that this approach fails, in general. According to (3.21a) and (3.22b), we get $\ddot{\boldsymbol{y}}(t) = -\boldsymbol{C}^\top \boldsymbol{\lambda}$ and $\boldsymbol{C}\ddot{\boldsymbol{y}}(t) = \ddot{\boldsymbol{z}}(t)$ resulting in Lagrange multipliers $\boldsymbol{\lambda}(t) = -(\boldsymbol{C}\boldsymbol{C}^\top)^{-1}\ddot{\boldsymbol{z}}(t)$. The BDF solution $\boldsymbol{x}_n = (\boldsymbol{y}_n, \boldsymbol{v}_n, \boldsymbol{\lambda}_n)^\top$ satisfies (3.13) and hence also

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}_n, t_n) = \boldsymbol{C}\boldsymbol{y}_n - \boldsymbol{z}(t_n), \ (n > 0)$$

since the constraints (3.21b) are part of the residual $\boldsymbol{F}$ in (3.33). A simple computation shows that the application of backward Euler method to (3.33), that is, BDF (3.13) with $k_n = 1$, yields

$$\boldsymbol{\lambda}_{n+1} = \frac{h_n + h_{n-1}}{2h_n}\boldsymbol{\lambda}(t_{n+1}) + \mathcal{O}(h_n) + \mathcal{O}\left(\frac{h_{n-1}^2}{h_n}\right).$$

For varying stepsizes, the numerical solution $\boldsymbol{\lambda}_{n+1}$ is completely wrong since it does not converge to the analytical solution $\boldsymbol{\lambda}(t_{n+1})$ if $h_n \to 0$ and h_{n-1} is fixed (see also [36]).

For some higher-order methods, convergence can still be guaranteed [23] but the errors in the numerical solution are typically substantially larger than in the ODE case. This is illustrated by simulation results for the hunting motion of the rigid wheelset in Figure 3.2. In this application, the BDF

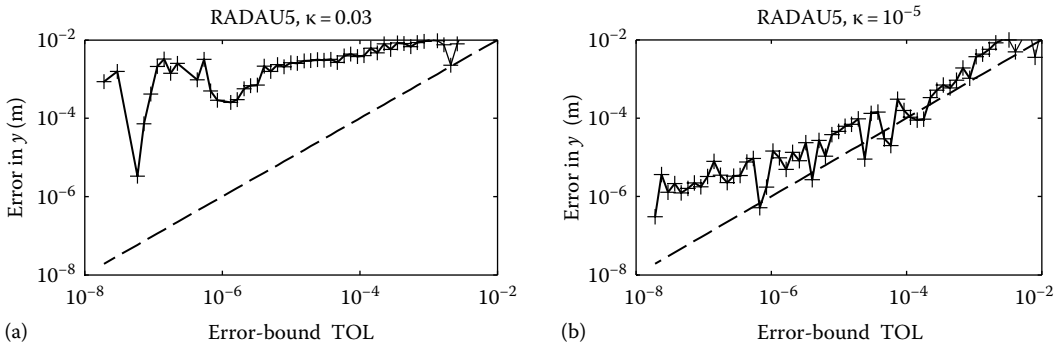


FIGURE 3.3 ODE solver **RADAU5** applied to the differential-algebraic equations of motion (3.21) of a rigid wheelset. (a) Solver parameter $\kappa = 0.03$, (b) Solver parameter $\kappa = 10^{-5}$.

integrator **DASSL** applied to (3.27) fails completely. As an alternative, the implicit Runge–Kutta solver **RADAU5** was used in the numerical tests [23]. For **RADAU5**, experience shows that the error of the numerical solution in ODE applications is usually kept well below the user-defined error tolerances TOL. But in the application to DAE (3.27), the relative error remains even for extremely small error bounds always in the size of 0.1%–1.0%. As a typical example, the left plot of Figure 3.3 shows the error in the lateral displacement y for various values of TOL.

To analyze these unsatisfactory results of **RADAU5,** we modify one of the internal parameters of the solver. BDF (3.13) and implicit Runge–Kutta methods define $\boldsymbol{x}_{n+1}$ as solution of a system of nonlinear equations. In the practical implementation, this system is solved iteratively by Newton's method that is stopped if the iteration error is less than $\kappa \cdot \mathrm{TOL}$. In this stopping criterion, the user-defined error tolerance for time integration (TOL) is scaled by a constant $\kappa \leq 1$ that is a free control parameter of the solver. Default values are $\kappa = 0.33$ in **DASSL** [24] and $\kappa = 0.03$ in **RADAU5** [23].

The right plot of Figure 3.3 shows that the error in time integration is reduced drastically and remains now roughly below the error bounds if κ is set to the very small value $\kappa = 10^{-5}$. The comparison of left and right plot in Figure 3.3 illustrates that the direct application of **RADAU5** to DAEs (3.27) makes the solver very sensitive w.r.t. (small) iteration errors in Newton's method.

An analytical perturbation analysis for (3.33) shows that the errors in $\boldsymbol{y}_n$, $\boldsymbol{v}_n$, and $\boldsymbol{\lambda}_n$ are of size $\mathcal{O}(\theta) + \mathcal{O}(\theta^2/h^2)$, $\mathcal{O}(\theta/h)$, and $\mathcal{O}(\theta/h^2)$, respectively, if the errors in the constraints (3.21b) are bounded by some small positive constant $\theta \ll 1$ (see [38]). Note that these error bounds *grow* unboundedly for decreasing time stepsize $h := \min_m h_m$.

This error amplification is typical of the direct application of ODE solvers to DAE (3.33). It may be considered as discrete analog of corresponding error bounds $\mathcal{O}(\max_t \|\boldsymbol{\theta}(t)\|) + \mathcal{O}(\max_t \|\dot{\boldsymbol{\theta}}(t)\|^2)$, $\mathcal{O}(\max_t \|\dot{\boldsymbol{\theta}}(t)\|)$, and $\mathcal{O}(\max_t \|\ddot{\boldsymbol{\theta}}(t)\|)$ for the analytical solution $\boldsymbol{y}(t)$, $\boldsymbol{v}(t)$, and $\boldsymbol{\lambda}(t)$ of (3.33) with perturbed constraints

$$\boldsymbol{\theta}(t) = \boldsymbol{g}(\boldsymbol{y}(t), t)$$

and $\max_t \left\|\boldsymbol{\theta}(t)\right\| \ll 1$ (see [38]).

For perturbations like $\boldsymbol{\theta}(t) = \varepsilon \sin \omega t$ that oscillate with small amplitude $\varepsilon \ll 1$ and high frequency $\omega \gg 1$, the errors in the solution $(\boldsymbol{y}(t), \boldsymbol{v}(t), \boldsymbol{\lambda}(t))$ are much larger than the perturbation itself since $\|\dot{\boldsymbol{\theta}}(t)\| = \mathcal{O}(\omega\varepsilon)$, $\|\ddot{\boldsymbol{\theta}}(t)\| = \mathcal{O}(\omega^2\varepsilon)$. A mechanical interpretation of these results is the well-known fact that high-frequency kinematic excitations $\boldsymbol{\theta}(t)$ in loop closing joints may cause large joint and constraint forces in a multi-body system.

3.4.2 Index Reduction and Projection

In the direct application of ODE methods to DAE (3.33), the robustness of the solvers may be improved substantially by small values of κ that result in (very) small iteration errors in Newton's method (see the right plot of Figure 3.3). On the other hand, values of κ that are less than 10^{-3} increase the number of Newton steps per time step substantially and may slow down the ODE solver dramatically.

Instead of applying ODE solvers directly to (3.33), it proved to be much more advantageous to transform the equations of motion *analytically* before time integration. These analytical transformations are guided by the perturbation analysis. In (3.21), the Lagrange multipliers $\boldsymbol{\lambda}(t)$ are defined implicitly by the constraints (3.21b). Differentiating these constraints $\mathbf{0} = \boldsymbol{g}(\boldsymbol{y}(t), t)$ twice, the acceleration constraints (3.22b) are obtained that define together with $\ddot{\boldsymbol{y}}(t) = \boldsymbol{M}^{-1}(\boldsymbol{y})(\boldsymbol{f} - \boldsymbol{G}^{\top}\boldsymbol{\lambda})$ (see 3.21a) the system of linear equations

$$\mathbf{0} = -[\boldsymbol{G}\boldsymbol{M}^{-1}\boldsymbol{G}^{\top}](\boldsymbol{y}, t)\boldsymbol{\lambda} + [\boldsymbol{G}\boldsymbol{M}^{-1}\boldsymbol{f}](\boldsymbol{y}, \dot{\boldsymbol{y}}, t) + \boldsymbol{g}^{(II)}(\boldsymbol{y}, \dot{\boldsymbol{y}}, t) \tag{3.34}$$

for the Lagrange multipliers $\boldsymbol{\lambda}(t)$. The $r = 2$ differentiation steps introduce the large error terms $\mathcal{O}(\|\ddot{\boldsymbol{\theta}}(t)\|)$ and $\mathcal{O}(\theta/h^2)$ in the analytical and numerical solution, respectively. In DAE terminology, the system (3.21) forms a DAE of index $r + 1 = 3$ that is called the *index-3 formulation* of the equations of motion [24].

For consistent initial values satisfying (3.23), the solution of the equations of motion (3.21) remains unchanged if the constraints (3.21b) on position level are substituted by the constraints (3.22a) on velocity level since

$$\boldsymbol{g}(\boldsymbol{y}(t), t) = \underbrace{\boldsymbol{g}(\boldsymbol{y}_0, t_0)}_{=0, \text{see (3.23)}} + \int_{t_0}^{t} \underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{g}(\boldsymbol{y}(\tau), \tau)}_{=0, \text{see (3.22a)}} \mathrm{d}\tau = \mathbf{0}. \tag{3.35}$$

From the numerical viewpoint, the *index-2 formulation* (3.21a/3.22a) of the equations of motion is attractive because only $r = 1$ differentiation of the velocity constraints (3.22a) is necessary to get the system (3.34) of linear equations defining $\boldsymbol{\lambda}(t)$.

With a perturbation $\boldsymbol{\theta}(t)$ in the velocity constraints (3.22a), the error terms for this index-2 DAE are of size $\mathcal{O}(\max_t \|\boldsymbol{\theta}(t)\| \, \|\dot{\boldsymbol{\theta}}(t)\|)$ for $\boldsymbol{y}(t)$, $\boldsymbol{v}(t)$ and of size $\mathcal{O}(\max_t \|\dot{\boldsymbol{\theta}}(t)\|)$ for $\boldsymbol{\lambda}(t)$. The error bounds for the numerical solution are $\mathcal{O}(\theta) + \mathcal{O}(\theta^2/h)$ for $\boldsymbol{y}_n$, $\boldsymbol{v}_n$ and $\mathcal{O}(\theta/h)$ for $\boldsymbol{\lambda}_n$. They are smaller by the (small) factor h than the corresponding error bounds for the original index-3 DAE (3.21) (see [38]). Applying **RADAU5** with the standard setting $\kappa = 0.03$ to the wheelset example of Figure 3.2 in index-2 formulation (3.21a/3.22a), the error of the numerical solution is kept well below the user-defined error tolerances TOL.

In the same way, the *index-1 formulation* (3.21a/3.22b) of the equations of motion is obtained substituting (3.21b) by the constraints (3.22b) on acceleration level. The error bounds for $\boldsymbol{\lambda}(t)$ and $\boldsymbol{\lambda}_n$ are slightly improved to $\mathcal{O}\|\max_t \|\boldsymbol{\theta}(t)\|)$ and $\mathcal{O}(\theta)$, respectively. BDF and implicit Runge–Kutta solvers applied to the index-1 formulation are as robust and efficient as in the classical ODE case.

In DAE terminology, the substitution of the constraints (3.21b) by one of its time derivatives is called *index reduction*. Index reduction by differentiation improves the solver's robustness substantially but suffers in long-term simulations from the *drift-off effect* that is illustrated by Figure 3.4.

Because of (3.35), the *analytical* solution of the index-2 formulation satisfies the original constraints (3.21b) exactly for all $t \geq t_0$. In the *numerical* solution, the integrand $(\mathrm{d}\boldsymbol{g}/\mathrm{d}t)(\boldsymbol{y}(\tau), \tau)$ in (3.35) is still bounded by a small constant $\varepsilon > 0$ but it does *not* vanish identically. Therefore, the error in (3.21b) may increase linearly in time t:

$$\|\boldsymbol{g}(\boldsymbol{y}_n, t_n)\| \leq \|\boldsymbol{g}(\boldsymbol{y}_0, t_0)\| + \int_{t_0}^{t_n} \epsilon \, \mathrm{d}t = \epsilon \cdot (t_n - t_0); \tag{3.36}$$

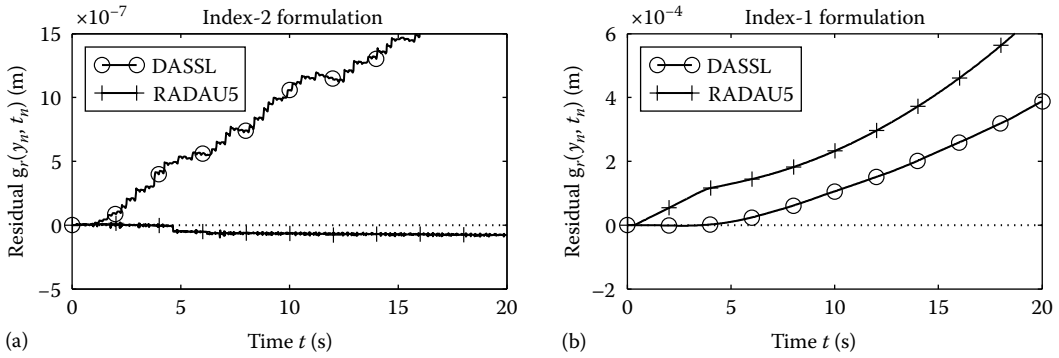


FIGURE 3.4 Drift-off effect in the dynamical simulation of the rigid wheelset of Figure 3.2 resulting in an increasing distance $g_r(\boldsymbol{y}, t)$ between the right wheel and the rail, i.e., in an increasing error in the constraints $\mathbf{0} = \boldsymbol{g} = (g_l, g_r)^\top$ that are defined by the contact conditions for left and right wheel. (a) Index-2 formulation, (b) Index-1 formulation.

the numerical solution $\boldsymbol{y}_n$ *drifts off* the manifold $\{(\boldsymbol{\eta}, t)\colon \boldsymbol{g}(\boldsymbol{\eta}, t) = \mathbf{0}\}$ that is defined by the constraints (3.21b) on position level. The error bound ε summarizes discretization and roundoff errors and the iteration errors of Newton's method.

For the index-1 formulation, a quadratic error growth $\|\boldsymbol{g}(\boldsymbol{y}_n, t_n)\| \leq \epsilon \cdot (t_n - t_0)^2$ has to be expected since the constraints (3.21b) on position level are substituted by their second derivatives (3.22b). Practical experience shows that ε depends on the solver and on the user-defined error tolerances TOL. In general, however, there is always a linear drift in the time integration of the index-2 formulation and a quadratic drift for the index-1 formulation (see Figure 3.4 for a typical example).

An early attempt to avoid both the numerical problems for the index-3 formulation (3.21) and the drift-off effect in the index-2 and index-1 formulation goes back to the work of Baumgarte [39] who substituted the constraints (3.21b) by a linear combination of all three constraints (3.21b, 3.22a, 3.22b). Because of the problems to select suitable coefficients for this linear combination (*Baumgarte coefficients*), the practical use of Baumgarte's approach is restricted to small-scale models (see also [40]).

Today, it is state of the art to avoid the drift-off effect by *projection* techniques [41,42]. During the time integration of the index-2 formulation, the residual $\|\boldsymbol{g}(\boldsymbol{y}_n, t_n)\|$ in the constraints (3.21b) is monitored. If the residual exceeds some user-defined small error bound $\epsilon > 0$, then $\boldsymbol{y}_n$ is projected onto the manifold $\{\boldsymbol{\eta}\colon \boldsymbol{g}(\boldsymbol{\eta}, t_n) = \mathbf{0}\}$ and the time integration is continued with the projected position coordinates $\hat{\boldsymbol{y}}_n$ instead of $\boldsymbol{y}_n$ (see Figure 3.5).

Mathematically, the projection defines a minimization problem

$$\|\boldsymbol{\eta} - \boldsymbol{y}_n\| \rightarrow \min_{\{\boldsymbol{\eta}\colon \boldsymbol{g}(\boldsymbol{\eta}, t_n) = \mathbf{0}\}} \tag{3.37}$$

that can be solved efficiently by simplified Newton iterations to get $\hat{\boldsymbol{y}}_n$ (see [42]).

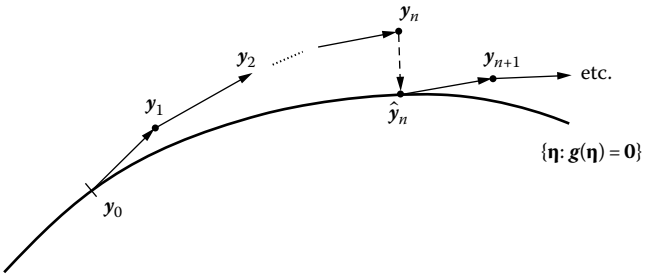


FIGURE 3.5 Time integration with projection steps for DAEs (3.21) with $\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{y})$.

For the index-1 formulation, the projection of both $\boldsymbol{y}_n$ and $\boldsymbol{v}_n$ helps to avoid the drift in the constraints (3.21b) and (3.22a) on position and velocity level [41].

3.4.3 Gear–Gupta–Leimkuhler Formulation

In the complex applications of vehicle system dynamics, the use of classical explicit projection methods like (3.37) is restricted to Runge–Kutta and other one-step methods since the efficient implementation in advanced BDF solvers with order and stepsize control is nontrivial.

Instead of implementing explicit projection steps in the solver, the equations of motion (3.21) are reformulated in a way that contains implicitly the projection onto the constraint manifold $\{\boldsymbol{\eta}: \boldsymbol{g}(\boldsymbol{\eta}, t_n) = \boldsymbol{0}\}$. The *Gear–Gupta–Leimkuhler formulation* (or *stabilized index-2 formulation*) of the equations of motion considers the constraints (3.21b) and (3.22a) on position and velocity level simultaneously [43]:

$$\begin{aligned} \dot{\boldsymbol{y}}(t) &= \boldsymbol{v} - \boldsymbol{G}^{\top}(\boldsymbol{y}, t)\boldsymbol{\mu}, \\ \boldsymbol{M}(\boldsymbol{y})\dot{\boldsymbol{v}}(t) &= \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{v}, t) - \boldsymbol{G}^{\top}(\boldsymbol{y}, t)\boldsymbol{\lambda}, \\ \boldsymbol{0} &= \boldsymbol{g}(\boldsymbol{y}, t), \\ \boldsymbol{0} &= \boldsymbol{G}(\boldsymbol{y}, t)\boldsymbol{v} + \boldsymbol{g}^{(I)}(\boldsymbol{y}, t). \end{aligned} \tag{3.38}$$

The increasing number of equations is compensated by a correction term $-\boldsymbol{G}^{\top}\boldsymbol{\mu}$ with auxiliary variables $\boldsymbol{\mu}(t) \in \mathbb{R}^{n_g}$. The correction term vanishes identically for the analytical solution ($\boldsymbol{\mu}(t) \equiv \boldsymbol{0}$) and remains in the size of the user-defined error tolerances TOL for the numerical solution.

Equation (3.38) forms an index-2 DAE that may be solved robustly and efficiently by BDF [43] and implicit Runge–Kutta methods [23]. However, the error estimates in classical ODE solvers tend to overestimate the local errors in the algebraic components $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ of DAE (3.38) (see [36]). Therefore, the components $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ should not be considered in the automatic stepsize control of BDF solvers [44]. For implicit Runge–Kutta solvers, the error estimates for $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are scaled by the small factor h_n.

The efficient evaluation of the correction term $\boldsymbol{G}^{\top}(\boldsymbol{y}, t)\boldsymbol{\mu}$ for given $\boldsymbol{\mu}, \boldsymbol{y}, t$ is nontrivial since multibody formalisms evaluate matrix-vector products $\boldsymbol{G}^{\top}\boldsymbol{\lambda}$ and $\boldsymbol{G}\boldsymbol{v}$ with O(N) complexity but do not evaluate the matrix $\boldsymbol{G}(\boldsymbol{y}, t)$ itself.

The key to the efficient solution of (3.38) was found in the analysis of a special class of overdetermined DAEs [45] that are closely related to DAEs (3.38) (see [46]). BDF solvers and implicit Runge–Kutta solvers for DAEs (3.12) approximate the Jacobian $(\alpha_{n,0}/h_n)(\partial \boldsymbol{F}/\partial \dot{\boldsymbol{x}}) + (\partial \boldsymbol{F}/\partial \boldsymbol{x})$ to compute $\boldsymbol{x}_{n+1}$ iteratively by Newton's method (see (3.10) and (3.14)). In (3.38), this Jacobian has a 4×4 block structure reflecting the four equations in (3.38) and the partitioning $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\lambda}, \boldsymbol{\mu})^{\top}$. Block (3,1) is an approximation of $(\partial \boldsymbol{g}/\partial \boldsymbol{y})(\boldsymbol{y}, t) = \boldsymbol{G}(\boldsymbol{y}, t)$, its transpose may be used to get the matrix-vector product $\boldsymbol{G}^{\top}\boldsymbol{\mu}$. This algorithm was implemented in the BDF solver **ODASSL** [45] that exploits furthermore the special structure of (3.38) to compute $\boldsymbol{\mu}_{n+1}$ more efficiently than in the standard **DASSL** implementation.

3.4.4 Advanced DAE Models

The numerical algorithms for DAE time integration that we discussed in detail for classical constrained mechanical systems with equations of motion (3.21) may be applied as well to the more complex hybrid DAEs (3.31/3.32) that are typical of vehicle system dynamics.

In the model equations in residual form (3.33), the vector $\boldsymbol{x}(t)$ is now given by $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{s}, \boldsymbol{v}, \boldsymbol{c}, \boldsymbol{\lambda}, \boldsymbol{w})^{\top}$. Time events and the discrete state equations (3.32) are handled by the methods that are known from ODE theory (see Section 3.2).

The differential equations (3.31b) and the algebraic equations (3.31c) and (3.31d) are added to the residual $\boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t)$. As a consequence, the BDF solution $\boldsymbol{x}_{n+1}$ of (3.31) satisfies the nonlinear equations (3.31d) exactly:

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}_{n+1}, \boldsymbol{s}_{n+1}, \boldsymbol{r}_j, t_{n+1}),$$

(see 3.13). The BDF solution $\boldsymbol{s}_{n+1}$ coincides with the function value $\boldsymbol{s}(\boldsymbol{y}_{n+1}, t_{n+1})$ of the function $\boldsymbol{s}(\boldsymbol{y}, t)$ that is implicitly defined by

$$\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}(\boldsymbol{y}, t), \boldsymbol{r}_j, t)$$

if $\partial \boldsymbol{h}/\partial \boldsymbol{s}$ is non-singular (see (3.25) and (3.31d)). BDF compute for a given geometric configuration $\boldsymbol{y}$ of the multi-body system the contact point coordinates $\boldsymbol{s}$ exactly up to the (small) iteration errors in Newton's method. The same result may be shown for the algebraic variables $\boldsymbol{w}$ in (3.31c).

With the regularity assumptions of Section 3.3, the model equations (3.31) form an index-3 DAE. Index reduction and projection techniques have to be applied. It is important to note that the nonlinear equations (3.31c) and (3.31d) may be solved directly w.r.t. the algebraic variables $\boldsymbol{w}$ and $\boldsymbol{s}$. In contrast to the constraints (3.31e), the nonlinear equations (3.31c) and (3.31d) do *not* imply hidden constraints on the state variables like (3.22).

The index reduction is based on time derivatives of the constraints $\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{s}, t)$. It does not involve any derivatives of the nonlinear equations $\boldsymbol{0} = \boldsymbol{b}(\boldsymbol{w}, \ldots)$ and $\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}, \boldsymbol{s}, t)$. The hidden constraints on the level of velocity coordinates are

$$\boldsymbol{0} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{g}(\boldsymbol{y}(t), \boldsymbol{s}(t), t) = \boldsymbol{G}(\boldsymbol{y}, \boldsymbol{s}, t)\boldsymbol{v} + \boldsymbol{g}^{(I)}(\boldsymbol{y}, \boldsymbol{s}, t)$$

with matrix $\boldsymbol{G}$ of (3.28) since $\dot{\boldsymbol{s}}(t)$ is a linear function of $\dot{\boldsymbol{y}}(t) = \boldsymbol{v}(t)$ that may be obtained differentiating $\boldsymbol{0} = \boldsymbol{h}(\boldsymbol{y}(t), \boldsymbol{s}(t), t)$ implicitly w.r.t. t (see also (3.29)).

Using a rather general concept of DAE index reduction [47], the Gear–Gupta–Leimkuhler formulation (3.38) is generalized to complex DAE models (3.31) (see [48,49]). Here, the Gear–Gupta–Leimkuhler formulation (3.38) is extended by (3.31b–d). To compensate the simultaneous consideration of the constraints on position level and the hidden constraints on velocity level, there is, as before, just one correction term. It has to be added to the kinematic equations $\dot{\boldsymbol{y}}(t) = \boldsymbol{v}$ that take the form

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{v} - \boldsymbol{G}^{\top}(\boldsymbol{y}, \boldsymbol{s}, t)\boldsymbol{\mu}$$

with the constraint matrix $\boldsymbol{G}$ of (3.28) and auxiliary variables $\boldsymbol{\mu}(t) \in \mathbb{R}^{n_g}$. Again, the correction term vanishes identically for the analytical solution and remains in the size of the user-defined error tolerances TOL for the numerical solution.

For this extended system, the Jacobian $\boldsymbol{J}$ in (3.14) has a 7×7 block structure with $(\partial \boldsymbol{g}/\partial \boldsymbol{y})$, $(\partial \boldsymbol{g}/\partial \boldsymbol{s})$, $(\partial \boldsymbol{h}/\partial \boldsymbol{s})$, $(\partial \boldsymbol{h}/\partial \boldsymbol{y})$ as blocks (6,1), (6,2), (5,2), and (5,1), respectively. In a modified version of the BDF solver **ODASSL**, these blocks are used for the efficient evaluation of the correction term $\boldsymbol{G}^{\top}\boldsymbol{\mu}$ with matrix $\boldsymbol{G}$ of (3.28) (see [48]).

3.4.5 Summary

The robust and efficient dynamical simulation in vehicle system dynamics may be based on DAE time integration methods with **DASSL** like BDF solvers that are applied to the Gear–Gupta–Leimkuhler formulation of the model equations. The approach is used successfully in industrial multi-body system simulation packages [4].

3.5 ALGORITHMS AND TOOLS FOR THE SIMULATION OF MULTI-PHYSICAL PROBLEMS

Typical applications of vehicle system dynamics are far beyond the field of classical multi-body analysis. Nevertheless, the methods and software tools of multi-body dynamics have been used very successfully as integration platform for large, complex models in vehicle system dynamics including mechanical, hydraulic, and electronic components.

Alternatives are general purpose simulation tools like Simulink® and modeling languages for complex physical systems like ACSL or Modelica.* Furthermore, the coupling of two or more mono-disciplinary simulation tools from different fields of application has proved to be very useful in industrial applications (*simulator coupling* or *co-simulation*).

3.5.1 Extensions of Multi-Body System Simulation Packages

In the early days of industrial multi-body system simulation, the focus was on specialized force elements with or without inner state variables $\boldsymbol{c}$ and $\boldsymbol{r}_j$ (see Section 3.3) [3].

Today, it might be even more important to embed the dynamical simulation of the mechanical system components efficiently in process chains for the simulation of the overall behavior of engineering systems. Figure 3.6 shows typical bidirectional interfaces between specialized simulation tools. The interfaces are used for data exchange in the pre- and post-processing of dynamical simulation.

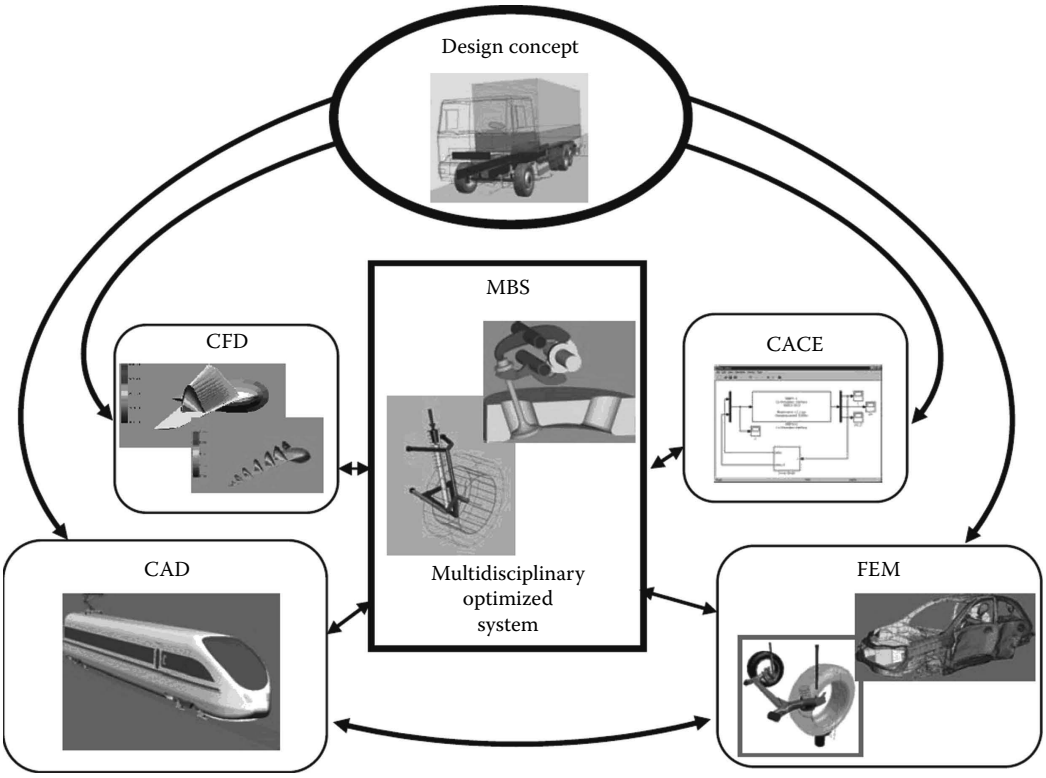


FIGURE 3.6 Integrated virtual design using multi-body system simulation software and its bidirectional interfaces. (From Veitl, A., *Integrierter Entwurf innovativer Stromabnehmer*, Fortschritt-Berichte VDI Reihe 12, Nr. 449, VDI-Verlag, Düsseldorf, Germany, 2001.)

* Simulink is a trademark of The MathWorks, Inc. ACSL is a trademark of The AEgis Technologies Group, Inc., and Modelica is a trademark of the Modelica Association.

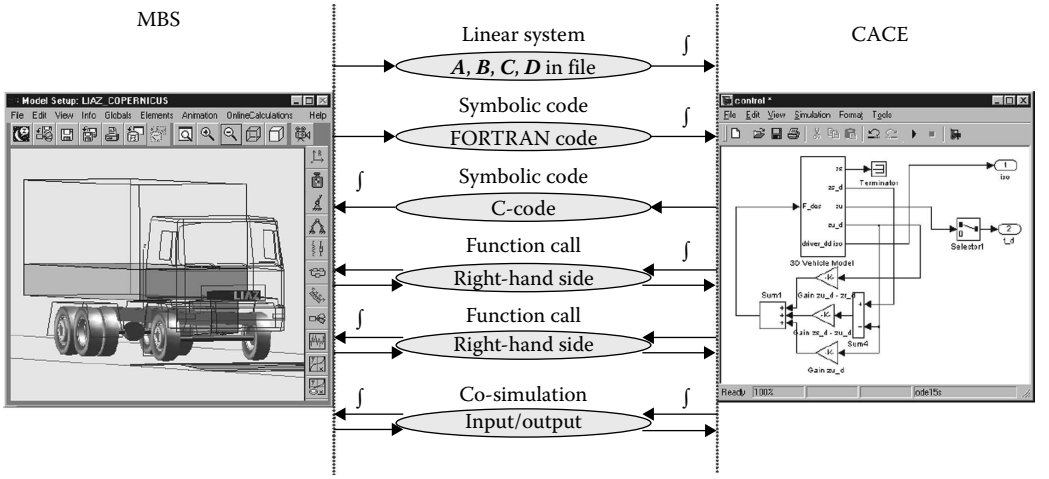


FIGURE 3.7 Typical interfaces between multi-body system tools and CACE tools. (From Vaculín, O. et al., *Vehicle Syst. Dyn*., 41:415–429, 2004.)

The simulation of flexible multi-body systems relies on the interfaces to finite element (FE) tools of structural dynamics. The interface from FE tool to multi-body system tool provides mass, damping, and stiffness data for the pre-processing of modes $\vec{w}^{(i)}$ in the Rayleigh–Ritz approach (3.16) (see [30]). The interface from multi-body system tool to FE tool supports the post-processing of multi-body system simulation data. Load vectors are transferred to the fatigue analysis in standard FE tools [51].

Modal reduction techniques are also used successfully in the analysis of certain aerodynamic effects. For this *close coupling* of computational fluid dynamics (CFD) and (flexible) multi-body dynamics, the classical interfaces between FE tools and multi-body system tools have been extended by an interface to CFD tools [16].

The interfaces between CAD and multi-body system tools support an efficient and fail-safe model setup using the geometrical data of CAD models. The opposite interface transfers all relevant geometrical modifications of the multi-body system model automatically back to the CAD model [52].

Controller synthesis is a central problem in many applications of vehicle system dynamics. Close links between the methods and tools for the dynamical simulation of mechanical components and the tools of computer-aided control engineering (CACE) are therefore essential. Figure 3.7 summarizes a number of typical interfaces between multi-body system software and CACE tools.

Already in the early 1990s, the increasing complexity of controllers in mechatronic system components motivated the integration of multi-body system models in CACE tools starting with the export of the system matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ describing the linearized equations of motion (see 3.6). Later, this interface was extended by the export of FORTRAN source code for nonlinear equations of motion (3.2). In both cases, the time integration for the full system including multi-body system model and controller is performed in the CACE tool.

Since the solvers of CACE tools are not tailored to the DAE model (3.31), it is often more attractive to perform the time integration of the full system including the controller in the multi-body system tool. State-of-the-art CACE tools support this approach by a code export interface. A typical example is the Real-Time Workshop®* of The MathWorks, Inc., that generates and executes standalone C code for developing and testing algorithms modeled in Simulink [53].

As an alternative to code export the tools of multi-body system analysis and CACE may be coupled as well by a function call interface that evaluates during time integration repeatedly the right-hand sides of the model equations for various given input data (see [53] for a more detailed discussion). The last row of Figure 3.7 refers to the co-simulation interface, which will be discussed later.

* Real-Time Workshop is a trademark of The MathWorks, Inc.

3.5.2 General Purpose Simulation Tools and Modeling Languages

Using interfaces between well-developed mono-disciplinary simulation tools, the experienced engineer may study multi-physical problems without leaving his usual software environment. Latest new developments of the powerful specialized tools are immediately available and the access to highly developed specialized solvers is straightforward.

Alternative approaches resolutely focus on the multi-domain aspect of complex multi-physical applications. The most widespread general purpose simulation tool from system theory is Simulink that has been developed as platform for multi-domain simulation and model-based design of dynamic systems. It has a block diagram user interface for model setup and a number of toolboxes for special fields of application like SimMechanics and SimDriveline.

The different concept behind specialized multi-body system tools on the one hand and general purpose simulation tools from system theory on the other is illustrated by the screenshots of user interfaces in Figure 3.7.

The increasing complexity of industrial multi-physical applications results in a revival of modeling languages that are designed to bundle the experience of specialists from various engineering domains. Modern modeling languages like Modelica [54] are object-oriented and do not follow the classical input–output scheme of traditional block-oriented tools (*noncausal* modeling).

Instead of classical ODE model equations, the model equations get the form of DAEs. Modelica is a rapidly developing nonproprietary modeling language that covers a wide range of applications like electrical circuits, vehicle dynamics, hydraulic and pneumatic systems, and automotive power trains (see [8] and the up-to-date material at the Modelica website [54] for more details of this ongoing development).

3.5.3 Simulator Coupling

Specialized mono-disciplinary tools and general purpose tools follow completely different strategies for the setup of multi-physical models and for the formulation of model equations. They have, however, in common that the full set of model equations is solved numerically in *one* simulation tool by *one* solver.

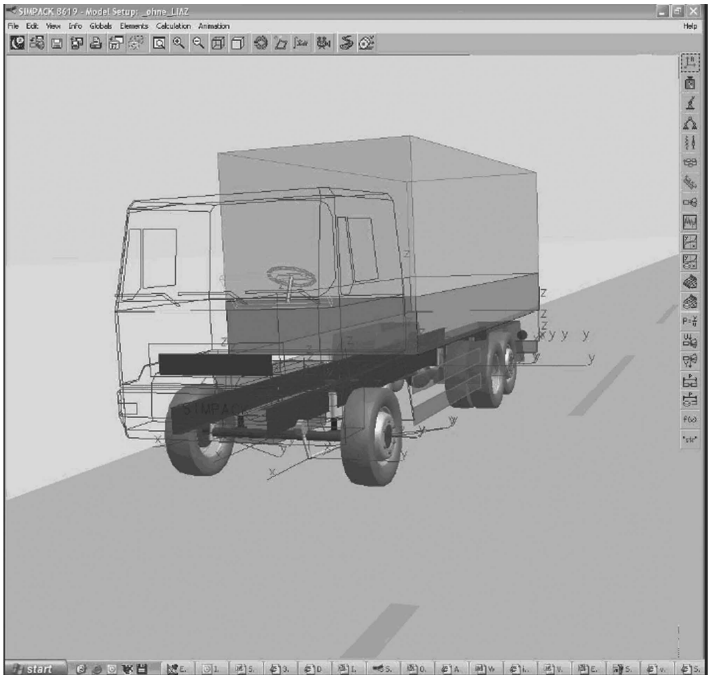
The modular structure of coupled multi-physical problems may also be exploited explicitly by coupling two or more well-established mono-disciplinary tools for model setup *and* for time integration (*simulator coupling* or *co-simulation*). In this way, the subsystems are handled by *different* solvers and each solver is tailored to the corresponding subsystem.

The communication between subsystems is restricted to discrete synchronization points T_j. For each subsystem, all necessary information from other subsystems has to be provided by interpolation or—if data for interpolation have not yet been computed—by extrapolation from $t \le T_j$ to the current *macro step* $T_j \to T_{j+1}$.

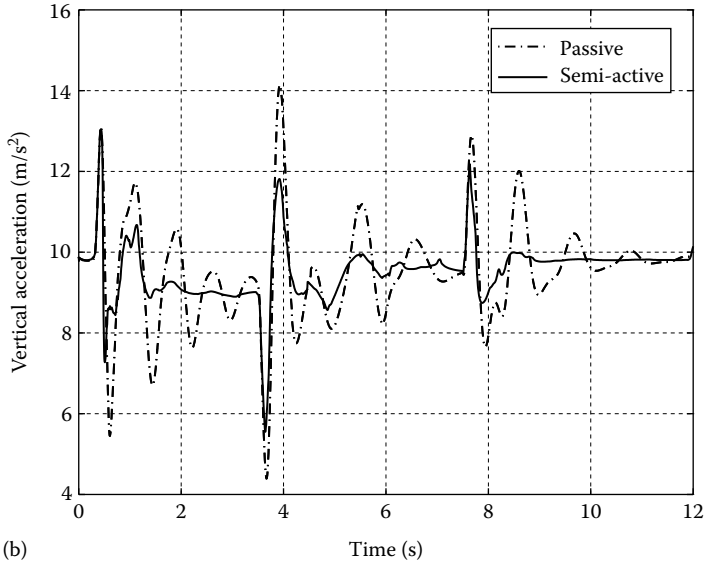
From the viewpoint of a multi-body system tool, the coupling variables to other simulation tools may be considered as a special type of discrete variables $\boldsymbol{r}_j$ in (3.32). In the multi-body system tool, the values of the coupling variables $\boldsymbol{r}_j$ are kept constant during the whole macro step $T_j \to T_{j+1}$. For all other simulation tools in the co-simulation environment, the update formula (3.32) involves the time integration from the synchronization point $t = T_j$ to the synchronization point $t = T_{j+1}$ to get $\boldsymbol{r}_{j+1} = \boldsymbol{k}(\ldots)$.

As a typical example of simulator coupling, Figure 3.8 shows simulation results for a heavy duty truck with semi-active suspensions [55]. The truck is modeled as a multi-body system with 41 bodies and 64 (mechanical) degrees of freedom. There are semi-active dampers in the axle suspension and in the cabin suspension that are controlled by an extended groundhook concept and by the skyhook law, respectively. Both controllers are modeled in Simulink and should be optimized to minimize dynamic road–tire forces and to maximize comfort.

In this application, the code export interfaces (see Figure 3.7) are not attractive since the multi-body system model contains closed loops resulting in constraints (3.31e) that cannot be handled by the standard solvers of Simulink. On the other hand, structure and parameters of the



(a)



(b)

FIGURE 3.8 Semi-active truck suspensions: vertical acceleration of the cabin (ISO filtered). (a): Multi-body system model, (b) Simulation data. (From Vaculín, O. et al., H. True, (ed.), *Proceedings of the 17th IAVSD Symposium on the Dynamics of Vehicles on Roads and on Tracks*, pp. 432–440. Supplement to *Vehicle Sys. Dyn.*, Vol. 37, Swets & Zeitlinger B.V., 2003.)

controllers are modified during the optimization process. It is therefore much more convenient to keep the controllers in Simulink instead of exporting their C code.

In [55], the problem was solved successfully by co-simulation of Simulink and the multi-body system tool SIMPACK [4]. The SIMPACK solver numerically integrates the mechanical part of the model, the control part is integrated in Simulink. The co-simulation sampling frequency is set to 200 Hz resulting in a constant *macro stepsize* $H = T_{j+1} - T_j = 5.0$ ms.

SIMPACK and Simulink run in two parallel processes. At the synchronization points, T_j the controller outputs, that is, the electrical signals defining the damper characteristics, are sent from Simulink to SIMPACK using inter-process communication (IPC). By the same IPC interface, Simulink gets the current vertical position, velocity and acceleration of truck chassis and cabin as controller input from SIMPACK. When the data exchange is completed, the solvers of Simulink and SIMPACK perform separately the macro step $T_j \rightarrow T_{j+1}$. Co-simulation continues with the data exchange at synchronization point T_{j+1} and so on.

The robust and reliable co-simulation contributed to the successful optimization of the controller design. A comparison of a truck with passive dampers and a truck with optimized semi-active dampers shows that the ride comfort is improved substantially (see the right plot of Figure 3.8 for the simulation of a typical driving maneuver [55]).

This case study illustrates that co-simulation techniques allow the convenient simulation and optimization of coupled problems in a classical software environment. Practical experience shows, however, that the coupling of different solvers may occasionally result in numerical instability. Furthermore, interpolation and extrapolation introduce additional discretization errors. The macro stepsize H has to be selected carefully. In typical standard applications, stability and accuracy may be guaranteed if H is in the range between 0.1 and 10 ms.

For certain classes of coupled problems, the instability phenomenon has been analyzed in great detail. Several modifications of the co-simulation techniques help to improve their stability, accuracy, and robustness also for larger macro stepsizes [56,57].

3.5.4 Summary

Complex multi-physical applications in vehicle system dynamics are beyond the field of multi-body dynamics. Successful simulation strategies use data interfaces or co-simulation interfaces between mono-disciplinary tools or models that are generated by general purpose simulation tools or modeling languages.

ACKNOWLEDGMENTS

The author is indebted to CISM International Centre for Mechanical Sciences, Udine (Italy), for the kind permission to include in the present chapter copyrighted material from the proceedings volume Simulation Techniques for Applied Dynamics (M. Arnold, W. Schiehlen eds., CISM Courses and Lectures, vol. 507, Springer, Wien New York, 2009). Furthermore, the author acknowledges gratefully the inspiring working atmosphere in late Professor W. Kortüm's Vehicle System Dynamics group at DLR German Aerospace Center and in the SIMPACK development team (1997–2002).

REFERENCES

1. W. Kortüm and W. Schiehlen. General purpose vehicle system dynamics software based on multibody formalisms. *Vehicle System Dynamics*, 14:229–263, 1985.
2. W. Kortüm and P. Lugner. *Systemdynamik und Regelung von Fahrzeugen*. Springer-Verlag, Berlin, Germany, 1994.
3. W. Kortüm, W.O. Schiehlen, and M. Arnold. Software tools: From multibody system analysis to vehicle system dynamics. In H. Aref and J.W. Phillips, (eds.), *Mechanics for a New Millennium*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2001, pp. 225–238.

4. W. Rulka. *Effiziente Simulation der Dynamik mechatronischer Systeme für industrielle Anwendungen*. PhD thesis, Department of Mechanical Engineering, Vienna University of Technology, Vienna, Austria, 1998.
5. E. Eich-Soellner and C. Führer. *Numerical Methods in Multibody Dynamics*. Teubner-Verlag, Stuttgart, Germany, 1998.
6. B. Simeon, C. Führer, and P. Rentrop. Differential-algebraic equations in vehicle system dynamics. *Surveys on Mathematics for Industry*, 1:1–37, 1991.
7. A.A. Shabana. *Computational Dynamics*, 2nd edn. John Wiley & Sons, Inc., New York, 2001.
8. M.M. Tiller. *Introduction to Physical Modeling with Modelica*. Kluwer Academic Publishers, Boston, MA, 2001.
9. R. Kübler and W. Schiehlen. Modular simulation in multibody system dynamics. *Multibody System Dynamics*, 4:107–127, 2000.
10. W.O. Schiehlen and P. Eberhard. *Technische Dynamik. Modelle für Regelung und Simulation*, 2nd edn. B.G. Teubner, Stuttgart, Germany, 2004.
11. H. Brandl, R. Johanni, and M. Otter. A very efficient algorithm for the simulation of robots and similar multibody systems without inversion of the mass matrix. In P. Kopacek, I. Troch, and K. Desoyer, (eds.), *Theory of Robots*. Pergamon Press, Oxford, U.K., 1988, pp. 95–100.
12. R.E. Roberson and R. Schwertassek. *Dynamics of Multibody Systems*. Springer-Verlag, Berlin, Germany, 1988.
13. A. Eichberger. *Simulation von Mehrkörpersystemen auf parallelen Rechnerarchitekturen*. Fortschritt-Berichte VDI Reihe 8, Nr. 332. VDI-Verlag, Düsseldorf, Germany, 1993.
14. C.T. Kelley. *Solving Nonlinear Equations with Newton's Method*. SIAM, Philadelphia, PA, 2003.
15. J.J. Moré, D.C. Sorensen, B.S. Garbow, and K.E. Hillstrom. The MINPACK project. In W.J. Cowell, (ed.), *Sources and Development of Mathematical Software*. Prentice-Hall, Englewood Cliffs, NJ, 1984, pp. 88–111.
16. W.R. Krüger and M. Spieck. Aeroelastic effects in multibody dynamics. *Vehicle System Dynamics*, 41:383–399, 2004.
17. C.T. Kelley. *Iterative Methods for Optimization*. SIAM, Philadelphia, PA, 1999.
18. T.F. Coleman, B.S. Garbow, and J.J. Moré. Software for estimating sparse Jacobian matrices. *ACM Transactions on Mathematical Software*, 10:329–345, 1984.
19. M. Arnold, A. Fuchs, and C. Führer. Efficient corrector iteration for DAE time integration in multibody dynamics. *Computer Methods in Applied Mechanics and Engineering*, 195:6958–6973, 2006.
20. U. Ascher and L.R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM, Philadelphia, PA, 1998.
21. E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations. I. Nonstiff Problems*. Springer-Verlag, Berlin, Germany, 1993.
22. C.B. Moler. *Numerical Computing with MATLAB*. SIAM, Philadelphia, PA, 2004.
23. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations. II. Stiff and Differential-Algebraic Problems*, 2nd edn. Springer-Verlag, Berlin, Germany, 1996.
24. K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical Solution of Initial–Value Problems in Differential-Algebraic Equations*, 2nd edn. SIAM, Philadelphia, PA, 1996.
25. A.A. Shabana. *Dynamics of Multibody Systems*, 2nd edn. Cambridge University Press, Cambridge, 1998.
26. O. Wallrapp. Linearized flexible multibody dynamics including geometric stiffening effects. *Mechanics of Structures and Machines*, 19:385–409, 1991.
27. R. Schwertassek and O. Wallrapp. *Dynamik flexibler Mehrkörpersysteme*. Vieweg, Braunschweig, 1999.
28. S. Dietz. *Vibration and Fatigue Analysis of Vehicle Systems using Component Modes*. Fortschritt-Berichte VDI Reihe 12, Nr. 401. VDI-Verlag, Düsseldorf, Germany, 1999.
29. B. Simeon. *Numerische Simulation gekoppelter Systeme von partiellen und differential-algebraischen Gleichungen in der Mehrkörperdynamik*. Fortschritt-Berichte VDI Reihe 20, Nr. 325. VDI-Verlag, Düsseldorf, Germany, 2000.
30. O. Wallrapp. Standardization of flexible body modeling in multibody system codes, Part I: Definition of standard input data. *Mechanics of Structures and Machines*, 22:283–304, 1994.
31. A. Heckmann, M. Arnold, and O. Vaculín. A modal multifield approach for an extended flexible body description in multibody dynamics. *Multibody System Dynamics*, 13:299–322, 2005.
32. R. von Schwerin. *Multibody System Simulation—Numerical Methods, Algorithms, and Software*, volume 7 of *Lecture Notes in Computational Science and Engineering*. Springer, Berlin, Germany, 1999.
33. F. Pfeiffer and Ch. Glocker. *Multibody Dynamics with Unilateral Contacts*. Wiley & Sons, New York, 1996.
34. R.A. Wehage and E.J. Haug. Generalized coordinate partitioning for dimension reduction in analysis of constrained dynamic systems. *Journal of Mechanical Design*, 104:247–255, 1982.
35. N. Orlandea. *Development and Application of Node-Analogous Sparsity-Oriented Methods for Simulation of Mechanical Dynamic Systems*. PhD thesis, University of Michigan, Ann Arbor, MI, 1973.

36. L.R. Petzold. Differential/algebraic equations are not ODEs. *SIAM Journal on Scientific and Statistical Computing*, 3:367–384, 1982.
37. M. Arnold and H. Netter. Wear profiles and the dynamical simulation of wheel-rail systems. In M. Brøns, M.P. Bendsøe, and M.P. Sørensen, (eds.), *Progress in Industrial Mathematics at ECMI 96*. Teubner, Stuttgart, Germany, 1997, pp. 77–84.
38. M. Arnold. A perturbation analysis for the dynamical simulation of mechanical multibody systems. *Applied Numerical Mathematics*, 18:37–56, 1995.
39. J. Baumgarte. Stabilization of constraints and integrals of motion in dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 1:1–16, 1972.
40. U.M. Ascher, H. Chin, and S. Reich. Stabilization of DAEs and invariant manifolds. *Numerische Mathematik*, 67:131–149, 1994.
41. E. Eich. Convergence results for a coordinate projection method applied to mechanical systems with algebraic constraints. *SIAM Journal on Numerical Analysis*, 30:1467–1482, 1993.
42. Ch. Lubich, Ch. Engstler, U. Nowak, and U. Pöhle. Numerical integration of constrained mechanical systems using MEXX. *Mechanics of Structures and Machines*, 23:473–495, 1995.
43. C.W. Gear, B. Leimkuhler, and G.K. Gupta. Automatic integration of Euler–Lagrange equations with constraints. *Journal of Computational and Applied Mathematics*, 12–13:77–90, 1985.
44. L.R. Petzold and P. Lötstedt. Numerical solution of nonlinear differential equations with algebraic constraints II: Practical implications. *SIAM Journal on Scientific and Statistical Computing*, 7:720–733, 1986.
45. C. Führer. Differential-algebraische Gleichungssysteme in mechanischen Mehrkörpersystemen. Theorie, numerische Ansätze und Anwendungen. PhD thesis, TU München, Mathematisches Institut und Institut für Informatik, 1988.
46. C. Führer and B. Leimkuhler. Numerical solution of differential-algebraic equations for constrained mechanical motion. *Numerische Mathematik*, 59:55–69, 1991.
47. P. Kunkel, V. Mehrmann, W. Rath, and J. Weickert. GELDA: A software package for the solution of general linear differential algebraic equations. *SIAM Journal on Scientific Computing*, 18:115–138, 1997.
48. M. Arnold. Numerical problems in the dynamical simulation of wheel-rail systems. *Zeitschrift für Angewandte Mathematik und Mechanik, Proceedings of ICIAM 95*, Issue 3:151–154, 1996.
49. M. Arnold, V. Mehrmann, and A. Steinbrecher. Index reduction in industrial multibody system simulation. Technical Report IB 532-01-01, DLR German Aerospace Center, Institute of Aeroelasticity, Vehicle System Dynamics Group, 2001.
50. A. Veitl. *Integrierter Entwurf innovativer Stromabnehmer*. Fortschritt-Berichte VDI Reihe 12, Nr. 449. VDI-Verlag, Düsseldorf, Germany, 2001.
51. S. Dietz, H. Netter, and D. Sachau. Fatigue life prediction by coupling finite-element and multibody systems calculations. In *Proceedings of DETC'97, ASME Design Engineering Technical Conferences DETC/VIB–4229*, Sacramento, CA, 1997.
52. W. Trautenberg. *Bidirektionale Kopplung zwischen CAD und Mehrkörpersimulationssystemen*. PhD thesis, Department of Mechanical Engineering, Munich University of Technology, Munich, Germany, 1999.
53. O. Vaculín, M. Valášek, and W.R. Krüger. Overview of coupling of multibody and control engineering tools. *Vehicle System Dynamics*, 41:415–429, 2004.
54. Modeling of Complex Physical Systems. Website of Modelica and the Modelica Association. http://www.modelica.org.
55. O. Vaculín, M. Valášek, and W. Kortüm. Multi-objective semi-active truck suspension by spatial decomposition. In H. True, (ed.), *Proceedings of the 17th IAVSD Symposium on the Dynamics of Vehicles on Roads and on Tracks*, pp. 432–440. Supplement to *Vehicle System Dynamics*, Vol. 37, Swets & Zeitlinger B.V., 2003.
56. R. Kübler and W. Schiehlen. Two methods of simulator coupling. *Mathematical and Computer Modelling of Dynamical Systems*, 6:93–113, 2000.
57. M. Arnold and M. Günther. Preconditioned dynamic iteration for coupled differential-algebraic systems. *BIT Numerical Mathematics*, 41:1–25, 2001.

4 Nonlinear Solid Mechanics with Finite Elements

Anna Pandolfi

CONTENTS

4.1 Introduction 70
4.2 Linear Problems 70
4.3 Nonlinear Problems 72
4.3.1 Solution Strategy 73
4.3.2 Solution Methods 75
4.4 Continuum Mechanics 75
4.4.1 Strong and Weak Forms of Equilibrium 76
4.4.2 Deformation Gradient 77
4.4.3 Deformation Tensors 78
4.4.4 Strain Rates 80
4.4.5 Strain Measures 83
4.4.6 Stress Measures 85
4.5 Nonlinear Formulations of the Virtual Work Equation 86
4.5.1 TL Formulation 87
4.5.2 UL Formulation 88
4.6 Constitutive Laws 90
4.6.1 Linear Elastic Behavior 91
4.6.2 Nonlinear Elastic Behavior 92
4.6.2.1 Incompressible Materials 93
4.6.3 Inelastic Behavior 95
4.6.3.1 Small Strain Elastoplasticity (von Mises) 96
4.6.3.2 Thermo-Elastoplasticity and Creep 101
4.6.4 Large Strain Elastoplasticity 102
4.6.4.1 Governing Equations 103
4.6.4.2 Solution Procedure 106
4.7 Finite Element Solution 108
4.7.1 Finite Element Linearization of the Equilibrium Equations 110
4.7.2 Matrix Equation for Solid Elements 111
4.7.2.1 Materially Nonlinear 112
4.7.2.2 TL Formulation 112
4.7.2.3 UL Formulation 113
4.8 Special Problems 114
4.8.1 Nonlinear Fracture Mechanics 114
4.8.2 Contact Mechanics 116
4.9 Conclusions 119
References 119

4.1 INTRODUCTION

A brief introduction to the finite element method in nonlinear mechanics is presented. The discussion focuses on purely mechanical problems excluding thermal processes. In view of finite element discretization, nonlinear theories of mechanics can be formulated in different ways, characterized by various advantages and difficulties. All such ways are here introduced and discussed.

The spatial discretization of the partial differential equations governing nonlinear problems leads to a system of nonlinear algebraic (for statics) or ordinary differential equations (for dynamics). Common problems in mechanics require to adopt nonlinear models. Nonlinearities are mainly due to the behavior of materials or to the magnitude of the displacements, or to the presence of unilateral constraints. Examples of such problems in the normal functioning of vehicles are the contact of the human subject with the vehicle, the mechanics of the tire, the mechanics of the bushings and of any rubber part in the vehicle, press fits, breaks pads, disks and calipers, and body structure. Examples of exceptional or pathological behaviors, that nevertheless have to be considered in the automotive design, are the mechanics of crash, the deployment of airbags, and the behavior of highly deformable parts designed to adsorb energy. The analysis of such problems can be pursued only with a nonlinear finite element approach.

The solution of a nonlinear system may be not unique, and the selection of a solver rather than another may privilege a particular solution and hide alternative solutions. Only the experience can help the analyst in selecting the most advantageous formulation and solver for a specific problem.

The following notes provide the indications on the basic tools necessary to analyze numerically the stress state of a generic mechanical model.

4.2 LINEAR PROBLEMS

The linearized theory of finite elements is based on several simplifying hypotheses concerning the behavior of solids and structures. In the literature, statics and dynamics problems able to satisfy such restrictive hypotheses are referred to as linear elasticity problems.

The linear elasticity assumptions are

1. The displacements are small. Under the action of external loads, bodies change their shape and their position in space. The small displacement hypothesis refers to a kinematics sufficiently small, such that
   a. The relative displacement between two close points is small. Thus, the displacement in a point neighborhood can be expressed as a linear function of the displacement of the point. The hypothesis does no exclude a priori the presence of large displacements or large rotations, but restricts the magnitude of the deformations. Thus, it allows writing the compatibility equations in linearized form. The linearized displacement field acquires the structure of a velocity field applied to the undeformed configuration (see Figure 4.1a).

      The expression of the derived small strains tensor ε_{ij} is

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right),$$

      where

      x_i is the coordinate of the particles of the solid

      $u_i = u_i(x_j)$ is the component of the displacement vector

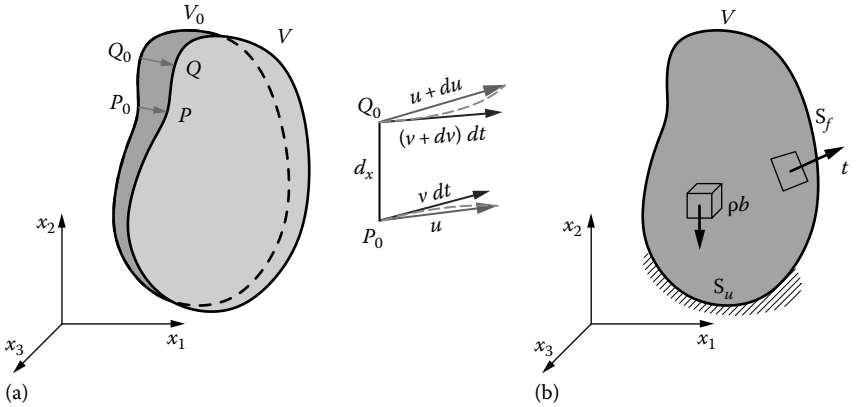


FIGURE 4.1 (a) Small displacement hypothesis. The displacements are equivalent to a velocity field applied to the undeformed configuration. (b) Static problem in small displacements. The equilibrium configuration reached under the external loads coincides with the undeformed configuration.

b. The displacements are small enough, so as not to alter the undeformed configuration of the body. In statics, the equilibrium condition is reached in the undeformed configuration (see Figure 4.1b). The hypothesis allows to write the equilibrium equations without accounting for the superposed kinematics and to decouple statics and kinematics:

$$\frac{\partial \sigma_{ij}}{\partial x_j} + \rho b_i = 0 \text{ in } V, \quad \sigma_{ij} n_i = t_j \text{ on } S_f,$$

where
σ_{ij} is the Cauchy stress tensor
ρ is the material density
b_i is the body forces per unit of mass
t_j is the surface tractions
n_i is the outward normal to the body external surface

2. The relationship between stress and strain is hyper-elastic and linear (Figure 4.2a). In an elastic material, the stress components are functions of the strain components. In a hyper-elastic material, a strain energy function acting as a potential for the stress is present. Hyper-elasticity induces reversibility, that is, a hyper-elastic material stores a certain amount of strain energy when loaded, and fully releases it when the load is removed. A linear constitutive law is characterized by linear relations between stress and strains. Linear elasticity automatically implies reversibility; therefore, the Hooke law is hyper-elastic.

   In linear hyper-elasticity, the relationship between Cauchy stress tensor σ_{ij} and small strain tensor ε_{ij} reads

$$\sigma_{ij} = C_{ijhk}\,\varepsilon_{hk}, \quad C_{ijhk} = \lambda \delta_{ij}\delta_{hk} + \mu(\delta_{ih}\delta_{jk} + \delta_{ik}\delta_{jh}), \tag{4.1}$$

where
δ_{ij} is the Kronecker delta
C_{ijhk} is the material stiffness tensor

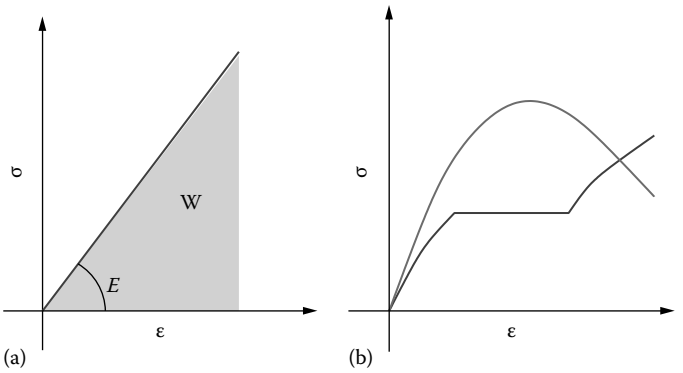


FIGURE 4.2 Material behaviors: (a) linear elasticity, implying reversibility and one-to-one correspondence and (b) nonlinear behavior, plasticity, softening, and in general nonreversibility.

The coefficients λ and μ are the Lamé constants:

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}, \quad \mu = G = \frac{E}{2(1+\nu)}.$$

The statement of a static or of a dynamic problem in linear elasticity is straightforward, because irreversible and time-dependent effects are ruled out and the solution is unique.

3. The boundary conditions do not change during the loading. The constrained part of the boundary remains constrained, and unilateral constraints are excluded.

Under a finite element discretization, an elastostatic problem reduces to a linear system of algebraic equations:

$$\boldsymbol{KU} = \boldsymbol{R} \tag{4.2}$$

where

U is the nodal displacement array
R is the external load array
K is the stiffness matrix

As a consequence of small strain and linear elasticity hypotheses, the stiffness matrix and the nodal force arrays are independent of the nodal displacements. Therefore, ***U*** is a linear function of ***R***. Obviously, Equation 4.2 holds only when the constraints on the displacement components do not change for any value of the external loads.

4.3 NONLINEAR PROBLEMS

When one of the previous assumptions is not satisfied, the problem is called nonlinear. Nonlinear problems can be classified according to the kind of nonlinearities considered. Table 4.1 collects the different types of nonlinear analyses [1].

When small displacements, small rotations, and infinitesimal strains are considered (SD-SR-SS), but the material behavior is nonlinear, the analysis is called materially nonlinear (MN). Stress and strain measures are the Cauchy tensor σ_{ij} and the small strain tensor ε_{ij}, respectively. The only nonlinearities accounted for are due to the stress–strain relationship (Figure 4.2b).

When finite displacements, finite rotations, small strains are considered (LD-LR-SS), and no special hypotheses are made on the nature of the material, it is assumed that the material undergoes

TABLE 4.1
Classification of Nonlinear Analyses

Analysis	Formulation	Stress Measure	Strain Measure
SD-SR-SS	MN	σ_{ij}	ε_{ij}
LD-LR-SS	TL	S_{ij}	E_{ij}
	UL	σ_{ij}	A_{ij}
LD-LR-LS	TL	S_{ij}	E_{ij}
	UL	σ_{ij}	ε_{ij}

small deformations in a reference system in motion with the body. In that case, it is possible to adopt two alternative formulations, both Lagrangian (i.e., the deformation process is described from a reference configuration).

The formulation is said Updated Lagrangian (UL) when the reference configuration is updated at every step of the evolution analysis. The stress and strain measures adopted are the Cauchy tensor σ_{ij} and the Almansi tensor A_{ij}, respectively.

The formulation is said Total Lagrangian (TL) when the reference configuration is always the same and coincides with the undeformed configuration. The stress and strain measures are the second Piola–Kirchhoff tensor S_{ij} and the right Green–Lagrange tensor E_{ij}, respectively.

The two formulations (UL and TL) can also be used in the large deformation, large rotation, and large displacement case (LD-LR-LS). In UL, the stress measure is the Cauchy tensor σ_{ij} and the strain measure is the Hencky tensor ε_{ij}. In TL, the stress measure is the second Piola–Kirchhoff tensor S_{ij} and the strain measure is the right Green–Lagrange tensor E_{ij}.

A totally different problem that may induce nonlinearities in the formulation is the change in the boundary conditions, as it happens in contact.

It is important to observe that the choice of a formulation is suggested by the problem at the hand. A preliminary accurate evaluation of the problem permits to select the correct formulation, avoiding painful fix-ups for pending ends or algorithmic inconsistencies. The most general formulation (TU) is always correct, but it can be expensive. Simplified formulations (MN, UL) may reduce the computational effort and provide a better understanding of the underling mechanics but they are not always accurate enough for the problem in consideration.

4.3.1 Solution Strategy

To solve a nonlinear mechanical problem, the equilibrium configurations of a deformable body under the action of external loads have to be found. A finite element discretization transforms the equilibrium condition in a nonlinear system of algebraic or ordinary differential equations, in the general form:

$$\boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t) = \boldsymbol{0}, \tag{4.3}$$

where

the array $\boldsymbol{R}(t)$ contains the external forces applied to the nodes

the array $\boldsymbol{F}^{\text{int}}(t)$ contains the internal forces, that is, the integral of the element stresses

Equation 4.3 includes all the nonlinearities of the problem. The external forces $\boldsymbol{R}(t)$ can be decomposed in the contributions of volume forces $\boldsymbol{R}_B(t)$, surface forces $\boldsymbol{R}_S(t)$, and contact forces $\boldsymbol{R}_C(t)$:

$$\boldsymbol{R}(t) = \boldsymbol{R}_B(t) + \boldsymbol{R}_S(t) + \boldsymbol{R}_C(t).$$

In a dynamical analysis, $\boldsymbol{R}(t)$ also includes inertia and damping forces.

The general expression for the internal forces $\boldsymbol{F}^{\text{int}}(t)$ is

$$\boldsymbol{F}^{\text{int}}(t) = \sum_{e} \int_{V^{(e)}(t)} \boldsymbol{B}^{(e)T} \boldsymbol{\sigma}^{(e)}(t)\, dV^{(e)},$$

where

$\boldsymbol{B}$ is the matrix that collects the strain-displacement functions
the index e runs over the element number

In general, the stress $\sigma^{(e)}(t)$ and the element volume $V^{(e)}(t)$ at time t are unknown.

Equation 4.3 has to be satisfied at every t time of the entire loading history, that is, for $0 \le t \le T$. In a static analysis with no viscous effects, the load level is the only relevant parameter; therefore, the time has the meaning of an ordering parameter. If time is used only to describe the sequence of load application, a dynamical analysis becomes equivalent to a static analysis, where the inertia effects are included.

In some cases, only the stress and the displacement fields in particular equilibrium configurations are of interest. Ad hoc methods exist able to identify such equilibrium configurations, avoiding the calculation of transient configurations. Of course, this can be done only if time-dependent behaviors are not important, and path-dependent geometrical and material nonlinearities are excluded.

In general cases, though, it is necessary to carry out an incremental step-by-step analysis. In an incremental step-by-step analysis, the event duration (or pseudo-duration) T is subdivided into time steps $\Delta t = t - t_0$ suitably chosen. The solution at time t_0 is known, while the solution at the time t has to be computed.

The nodal forces $\boldsymbol{F}^{\text{int}}(t)$ can be expressed as the sum of the previous step solution and of the increment $\Delta \boldsymbol{F}$, which accounts for the displacement and stress changes in the interval Δt:

$$\boldsymbol{F}^{\text{int}}(t) = \boldsymbol{F}^{\text{int}}(t_0) + \Delta \boldsymbol{F}. \tag{4.4}$$

The array $\Delta \boldsymbol{F}$ can be rendered in approximate form through a linearization, which introduces the stiffness matrix $\boldsymbol{K}(t_0)$ at time t_0:

$$\Delta \boldsymbol{F} = \boldsymbol{K}(t_0)\, \Delta \boldsymbol{U}, \quad \boldsymbol{K}(t_0) = \left. \frac{\partial \boldsymbol{F}^{\text{int}}}{\partial \boldsymbol{U}} \right|_{t=t_0}. \tag{4.5}$$

$\boldsymbol{K}(t_0)$ is the derivative of the nodal forces $\boldsymbol{F}^{\text{int}}$ with respect to the nodal displacements $\boldsymbol{U}$ at time t_0. The combination of 4.4, 4.5, and 4.3 gives

$$\boldsymbol{K}(t_0)\, \Delta \boldsymbol{U} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0), \quad \Delta \boldsymbol{U} = \boldsymbol{K}(t_0)^{-1} [\boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0)]. \tag{4.6}$$

The solution of 4.6 is an approximation of the incremental nodal displacements in the interval Δt:

$$\boldsymbol{U}(t) = \boldsymbol{U}(t_0) + \Delta \boldsymbol{U}. \tag{4.7}$$

For an accurate evaluation of stresses and forces, the nonlinear system (4.3) requires an iterative procedure. A rough approximation of the solution may lead to numerical instabilities.

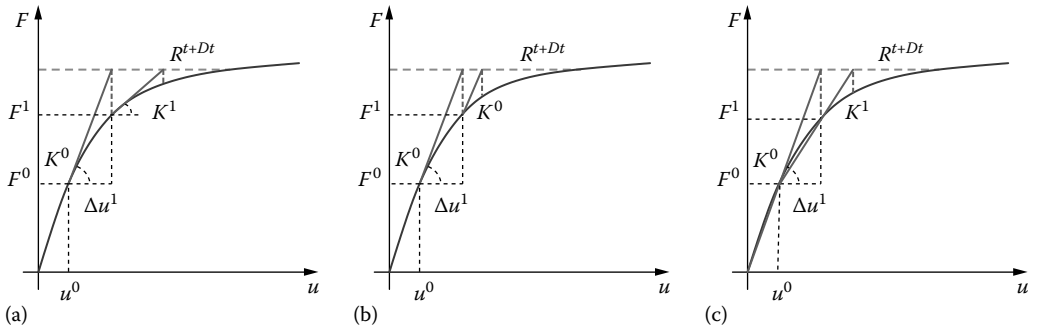


FIGURE 4.3 Schematic of (a) Newton–Raphson iterative procedure. (b) Modified Newton–Raphson iterative procedure. (c) Secant iterative procedure.

4.3.2 Solution Methods

The most widely used algorithm for the solution of nonlinear system is the Newton–Raphson method (Figure 4.3a). Equations 4.6 and 4.7 are applied iteratively. The solution of iteration k is used as the starting point for the next iteration $(k + 1)$:

$$\begin{aligned} \Delta \boldsymbol{U}^{k+1} &= \boldsymbol{K}^{k-1}[\boldsymbol{R}(t) - \boldsymbol{F}^k] \\ \boldsymbol{U}^{k+1} &= \boldsymbol{U}^k + \Delta \boldsymbol{U}^{k+1}. \end{aligned} \tag{4.8}$$

Stresses, nodal forces $\boldsymbol{F}^{k+1}$, and tangent stiffness $\boldsymbol{K}^{k+1}$ at iteration $(k + 1)$ are computed through the computed displacements (4.8). The error $\boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t)^k$ is called "residual." The residual is an array of out-of-balance forces that requires a further increment of the displacements to be equilibrated. The Newton–Raphson method is very efficient, but it may be excessively demanding in terms of computational cost. In fact, at each iteration it requires the evaluation of the nodal forces array and of the tangent stiffness matrix.

Alternative methods can be used to reduce the computational effort. In the modified Newton (or quasi-Newton) method, the tangent stiffness is computed only at the beginning of the load step, and is considered constant during the iterative procedure (see Figure 4.3b). In the secant method, the tangent stiffness is approximated with the secant stiffness. Its calculation requires the knowledge of the nodal forces and displacements at two sequential iterations, but it does not need the derivation of the analytical expression of the stiffness matrix (see Figure 4.3c).

The iterative procedure terminates under the satisfaction of a convergence test. In general, the convergence test is based on the norm of the arrays involved in the solution, such as the residual norm (or the incremental displacement norm). The solution is considered correct when the global array norm is smaller than a small percentage of the norm of the residuals at the iteration zero (or of the norm or the total displacement):

$$|\boldsymbol{R}^{k+1}| \leq \epsilon\, |\boldsymbol{F}^{\text{ext}}|, \quad |\Delta \boldsymbol{U}^{k+1}| \leq \epsilon\, |\boldsymbol{U}|,$$

where the small constant ε (e.g., 10^{-12}) is the desired tolerance.

4.4 CONTINUUM MECHANICS

The description of the nonlinear finite element approaches requires a brief overview of the continuum mechanics variables and equations.

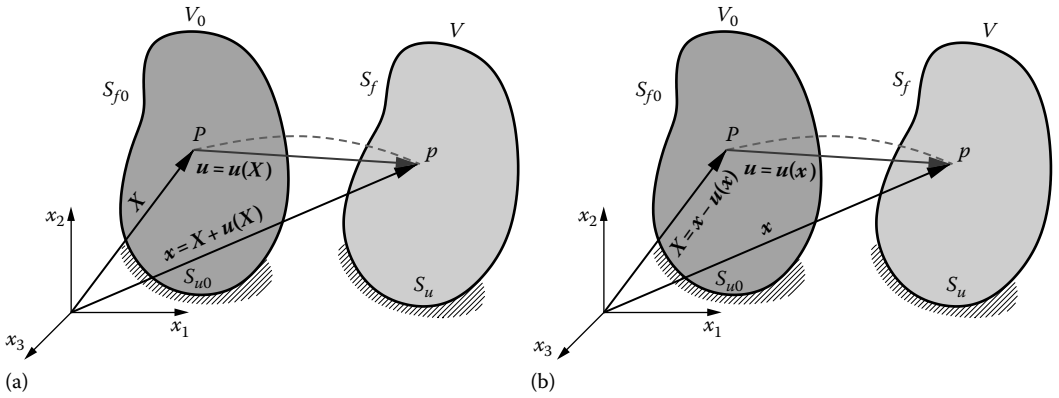


FIGURE 4.4 A continuum in the initial and deformed configuration, definition of the variables. (a) Lagrangian approach. (b) Eulerian approach.

A body, subject to constraints, surface tractions, and body forces (all variables in time), undergoes a motion described by a regular mapping φ. The position of the body is identified by a fixed Cartesian reference system. The equilibrium configurations of the body at regular time intervals Δt have to be determined.

The motion of a continuum without restrictions on the magnitude of displacements can be studied with two alternative point of views:

- In a Lagrangian or "material" formulation, one tracks the evolution of all the body particles from the initial configuration to the final configuration (Figure 4.4a).
- In an Eulerian or "spatial" formulation, one follows the motion of the material through a reference volume (Figure 4.4b). The Eulerian formulation is used widely in fluid mechanics.

In the following, only the Lagrangian description will be considered. A detailed derivation of the equations that govern a continuum mechanics problem in both descriptions can be found in several well-known books, for example, [2–4].

The reference variables are identified with a capital symbol, while the current configuration variables are identified with lowercase symbols. When an uppercase symbol is necessary for both configurations, a superscript 0 index will distinguish the reference configuration from the current, see, for example, the volume V^0 and V, respectively.

4.4.1 Strong and Weak Forms of Equilibrium

As in the linear case, the equilibrium of the continuum at time t is expressed in weak form through the virtual work theorem:

$$\int_V \sigma_{ij}\delta e_{ij}\, dV = W(t), \quad \delta e_{ij} = \frac{1}{2}\left(\frac{\partial \delta u_i}{\partial x_j} + \frac{\partial \delta u_j}{\partial x_i}\right), \tag{4.9}$$

where

x_i is the spatial coordinate of a point at time t

δu is the component of a virtual (i.e., compatible) displacement field added to the deformed configuration at time t

σ_{ij} is the Cauchy stress tensor, expressing the forces per unit of area in the current configuration

δe_{ij} is the symmetric part of the displacement gradient, derived from the virtual displacements

V is the body volume at time t

The right-side member of Equation 4.9_1 represents the virtual external work:

$$W(t) = \int_V \rho b_i \delta u_i \, dV + \int_{S_f} t_i \delta u_i \, dS, \tag{4.10}$$

where

ρb_i is the body force per unit of volume at time t
t_i is the surface traction at time t
S_f defines the part of the external boundary where tractions are applied

Equation 4.9_1 is derived by considering an infinitesimal variation of the current equilibrium configuration. It corresponds to the strong equilibrium equations in the volume V:

$$\frac{\partial \sigma_{ij}}{\partial x_j} + \rho b_i = 0 \tag{4.11}$$

and on the boundary S_f

$$\sigma_{ij} n_i = t_j, \tag{4.12}$$

where n_i defines the outward normal on S_f at time t.

It is important to observe that

1. The virtual work equation is obtained from the equilibrium equations (4.11 and 4.12), transformed from vectorial equation to a scalar through a virtual displacement field, and integrated over continuum volume and surface.
2. The components of the virtual deformation tensor in $(4.9)_2$ are obtained, like in the small strain case, by differentiation of the displacement field with respect to the current coordinates x_i at time t. The expression of the virtual deformation does not depend on the magnitude of the actual displacements.
3. The virtual displacements δu_i are a compatible first variation of the true displacements u_i. Compatibility implies satisfaction of the displacement conditions on the constrained boundary and of the internal compatibility constraints.
4. Formally, the virtual displacement equation is identical to the one used in linear analyses, except that it is written in the current configuration. Therefore, all the results obtained for the linear virtual work equation hold also for Equation 4.9_1.
5. The substantial difference with respect to linear analyses is that the current configuration at time t is unknown. This has strong implications on solution procedures. For example, it is not possible to apply the principle of the superposition of effects; also, in the evaluation of the Cauchy stress, it is necessary to account for the presence of rigid rotations.
6. To follow the configuration changes of a body in finite deformations, it is necessary to use work conjugate stress and strain measures, related through constitutive equations.

4.4.2 Deformation Gradient

A deformation mapping φ modifies the configuration of a continuum. In a fixed reference system, the position of a particle P in the initial or undeformed configuration is identified by the coordinates $\boldsymbol{X} = \{X_i\}$. The position of the same particle in the current or deformed configuration, at time t,

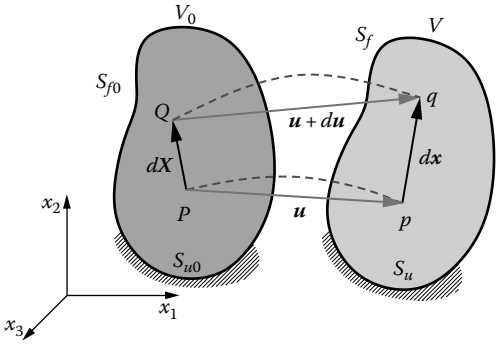


FIGURE 4.5 Configuration change in a continuum undergoing a mapping φ. The segment $d\boldsymbol{X}$ transforms into $d\boldsymbol{x} = \boldsymbol{F}d\boldsymbol{X}$.

is identified with $\boldsymbol{x} = \{x_i\}$. The current position is obtained by the initial position by adding the displacements to the initial position, $\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}$ (or in components $x_i = X_i + u_i$). For particles sufficiently close, the configuration change is described completely by the deformation gradient $\boldsymbol{F}$:

$$\boldsymbol{F} = \frac{d\boldsymbol{x}}{d\boldsymbol{X}} = \boldsymbol{I} + \frac{d\boldsymbol{u}}{d\boldsymbol{X}}, \quad F_{ij} = \frac{dx_i}{dX_j} = \delta_{ij} + \frac{du_i}{dX_j}.$$

$\boldsymbol{F}$ is the Jacobian matrix of the mapping φ and is a real value nonsymmetric tensor. The determinant $J = \det(\boldsymbol{F})$ is called the Jacobian of the transformation. Figure 4.5 shows that, under the mapping φ, the deformation gradient transforms the segment $d\boldsymbol{X}$ into the segment $d\boldsymbol{x} = \boldsymbol{F}d\boldsymbol{X}$.

If the deformation mapping is continuum and regular, J is nonzero and $\boldsymbol{F}$ is invertible:

$$\boldsymbol{F}^{-1} = \frac{d\boldsymbol{X}}{d\boldsymbol{x}} = \boldsymbol{I} - \frac{d\boldsymbol{u}}{d\boldsymbol{x}}, \quad F_{ij}^{-1} = \frac{dX_i}{dx_j} = \delta_{ij} - \frac{du_i}{dx_j}.$$

It is possible to prove that J measures the ratio between the local initial density ρ^0 and the local current density ρ:

$$J = \frac{\rho^0}{\rho} > 0.$$

Clearly, for physical reasons, negative values of J are not acceptable.

The deformation gradient $\boldsymbol{F}$ contains all the information about the configuration change in the neighborhood of a point, including deformations and rigid rotations. Theoretically, $\boldsymbol{F}$ could be used in the constitutive relations, although the rigid rotation part has to be removed. For practical reasons, however, it is more convenient to introduce measures of deformations that exclude rigid rotations.

4.4.3 Deformation Tensors

To eliminate rigid rotations from $\boldsymbol{F}$, it is possible to define symmetric measures of deformation.

By pre- or post-multiplication of $\boldsymbol{F}$ with the transpose $\boldsymbol{F}^T$, the right Cauchy–Green stretch tensor $\boldsymbol{C}$ and left Cauchy–Green stretch tensor $\boldsymbol{B}$, respectively, are obtained:

$$\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}, \quad \boldsymbol{B} = \boldsymbol{F}\boldsymbol{F}^T.$$

$\boldsymbol{C}$ is a symmetric real-value tensor, with three real eigenvalues, C_α, and three orthogonal eigenvectors, $\boldsymbol{N}_\alpha$. The eigenvectors define the three principal directions of stretch in the undeformed (or material) reference configuration. Also, $\boldsymbol{B}$ is a symmetric real-value tensor, and possesses three real eigenvalues B_α, and three orthogonal eigenvectors, $\boldsymbol{n}_\alpha$. The eigenvectors define three principal directions of stretch in the deformed or spatial reference configuration.

The right stretch tensor $\boldsymbol{U}$ and the left stretch tensor $\boldsymbol{V}$ (related to pure deformations) are defined as

$$\boldsymbol{C} = \boldsymbol{U}^2, \quad \boldsymbol{U} = \boldsymbol{C}^{1/2}, \quad \boldsymbol{B} = \boldsymbol{V}^2, \quad \boldsymbol{V} = \boldsymbol{B}^{1/2}.$$

$\boldsymbol{C}$ and $\boldsymbol{B}$ are definite positive tensors, and their eigenvalues C_α and B_α are positive. The three real numbers $\lambda_\alpha = \sqrt{C_\alpha}$ are called principal stretches. The tensors $\boldsymbol{C}$ and $\boldsymbol{B}$ admit the spectral decomposition:

$$\boldsymbol{C} = \sum_{\alpha=1}^{3} C_\alpha \boldsymbol{N}_\alpha \otimes \boldsymbol{N}_\alpha, \quad \boldsymbol{B} = \sum_{\alpha=1}^{3} B_\alpha \boldsymbol{n}_\alpha \otimes \boldsymbol{n}_\alpha.$$

Clearly, $\boldsymbol{U}$ and $\boldsymbol{V}$ also admit a spectral decomposition:

$$\boldsymbol{U} = \sum_{\alpha=1}^{3} \lambda_\alpha \boldsymbol{N}_\alpha \otimes \boldsymbol{N}_\alpha, \quad \boldsymbol{V} = \sum_{\alpha=1}^{3} \sqrt{B_\alpha}\, \boldsymbol{n}_\alpha \otimes \boldsymbol{n}_\alpha. \tag{4.13}$$

Alternative deformation tensors are obtained by decomposing the deformation gradient into the product of a deformation tensors and a rotation matrix $\boldsymbol{R}$, in the two alternative forms:

$$\boldsymbol{F} = \boldsymbol{R}\boldsymbol{U}, \quad \boldsymbol{F} = \boldsymbol{V}\boldsymbol{R}, \quad \boldsymbol{V} = \boldsymbol{R}\boldsymbol{U}\boldsymbol{R}^T. \tag{4.14}$$

Equation 4.14_1 defines the right polar decomposition of the deformation gradient, while Equation 4.14_2 defines the left polar decomposition. The rotation matrix $\boldsymbol{R}$ describes the rigid rotation of the deformed configuration with respect to the undeformed.

The meaning of the polar decomposition is that any deformation may be thought as the application of a pure deformation (right stretch) $\boldsymbol{U}$, described in the principal directions of the undeformed reference configuration, followed by a rigid rotation $\boldsymbol{R}$ $(4.14)_1$; or, alternatively, as the application of a rigid rotation followed $\boldsymbol{R}$ by a pure deformation (left stretch) $\boldsymbol{V}$ in the principal directions of the deformed configuration $(4.14)_2$. The two polar decompositions are visualized in Figure 4.6 in the case of a plane pure shear deformation. The components of the matrix $\boldsymbol{R}$ are defined by the rigid rotation angle β.

If $\boldsymbol{\Lambda}$ denotes the diagonal tensor that contains the principal stretches λ_α, and $\boldsymbol{R}_L$ the matrix that lists the three eigenvectors $\boldsymbol{N}_\alpha$, Equation 4.13_1 may be written as

$$\boldsymbol{U} = \boldsymbol{R}_L \boldsymbol{\Lambda} \boldsymbol{R}_L^T. \tag{4.15}$$

The components of the matrix $\boldsymbol{R}_L$ are defined by the angle φ illustrated in Figure 4.6, which identifies the position of the principal stretch angles before the application of the rigid rotation.

From $(4.14)_3$ and (4.15), one obtains

$$\boldsymbol{V} = \boldsymbol{R}_E \boldsymbol{\Lambda} \boldsymbol{R}_E^T, \quad \boldsymbol{R}_E = \boldsymbol{R}\boldsymbol{R}_L, \tag{4.16}$$

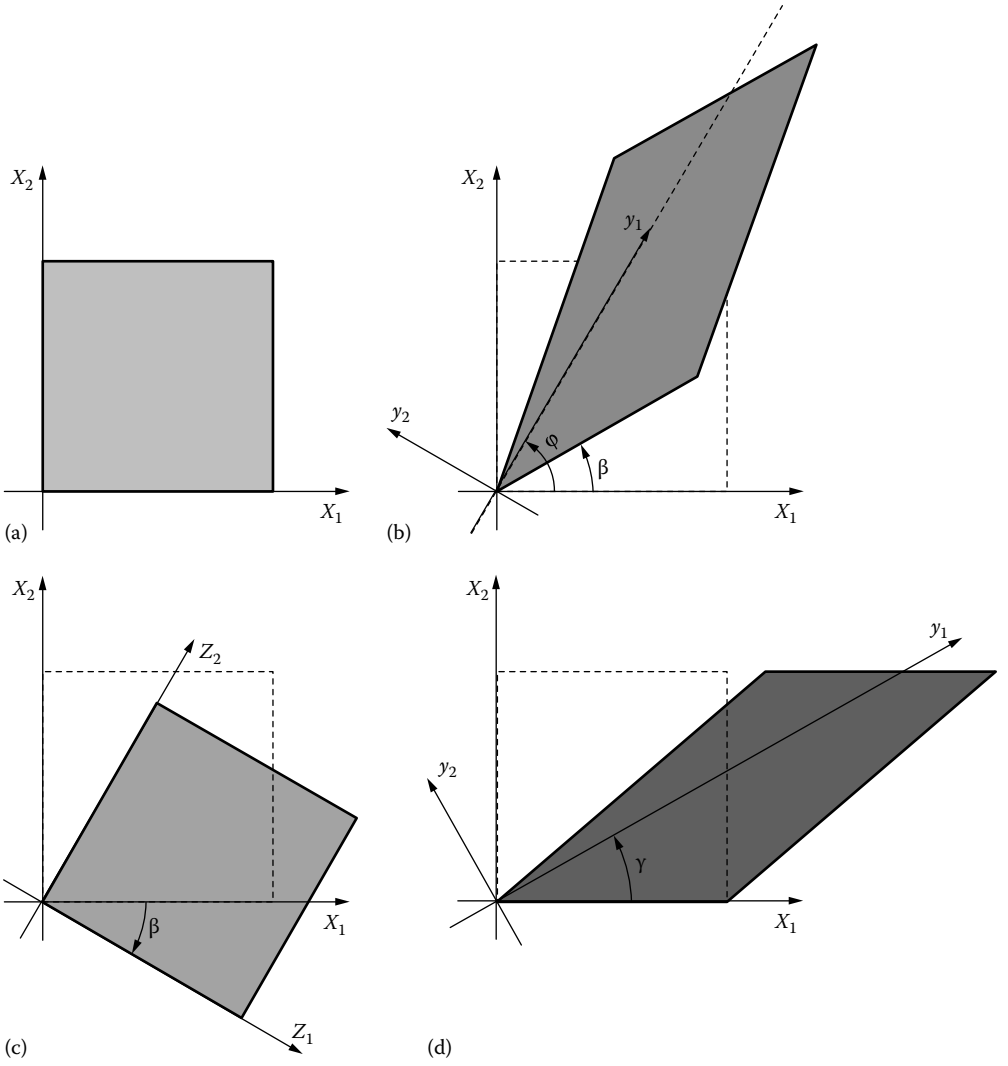


FIGURE 4.6 Visualization of the polar decomposition in the pure shear case. (a)→(b)→(d) right decomposition, (a)→(c)→(d) left decomposition.

where $\boldsymbol{R}_E$ collects the eigenvectors of $\boldsymbol{B}$. The components of the matrix $\boldsymbol{R}_E$ are defined by the angle γ, illustrated in Figure 4.6, as the angle that identifies the position of the principal stretch axes. By comparing $(4.13)_2$ and (4.16), it is clear that

$$\sqrt{B_\alpha} = \lambda_\alpha, \quad \boldsymbol{n}_\alpha = \boldsymbol{R}\boldsymbol{N}_\alpha,$$

that is, the eigenvectors of $\boldsymbol{C}$ and $\boldsymbol{B}$ are equal, while their eigenvectors are rotated by the angle $\boldsymbol{R}$. For the shear example in Figure 4.6, this means that $\varphi = \beta + \gamma$.

4.4.4 Strain Rates

In the description of the motion of a material particle, it is necessary to define the rates of the deformation measures previously mentioned [5,6]. A few definitions are introduced in the following.

The rotation rate $\dot{\boldsymbol{R}}$ is given by

$$\dot{\boldsymbol{R}} = \boldsymbol{\Omega}_R \boldsymbol{R}, \tag{4.17}$$

where $\boldsymbol{\Omega}_R$ is the skew-symmetric spin tensor. It is evident that at the initial time (when deformed and undeformed configurations coincide) $\boldsymbol{R} = \boldsymbol{I}$, and this helps in understanding the meaning of $\boldsymbol{\Omega}_R$. The rotations of the principal reference systems referred as the undeformed and deformed configuration, respectively, are expressed by analogous relations:

$$\dot{\boldsymbol{R}}_L = \boldsymbol{R}_L \boldsymbol{\Omega}_L, \quad \dot{\boldsymbol{R}}_E = \boldsymbol{R}_E \boldsymbol{\Omega}_E$$

The tensors $\boldsymbol{\Omega}_L$ and $\boldsymbol{\Omega}_E$ are called spin of the principal reference systems in the undeformed and deformed configuration.

The velocity gradient $\boldsymbol{L}$ is defined as

$$\boldsymbol{L} = \frac{d\boldsymbol{v}}{d\boldsymbol{x}}, \quad \boldsymbol{v} = \frac{d\boldsymbol{x}}{dt} = \frac{d\boldsymbol{u}}{dt},$$

where $\boldsymbol{v}$ is the velocity of a particle in the deformed configuration (Figure 4.7). The tensor $\boldsymbol{L}$ has the typical linear structure of a velocity field, and decomposes into the sum of two tensors (Figure 4.8): the symmetric deformation rate tensor $\boldsymbol{D}$ and the skew-symmetric spin tensor $\boldsymbol{W}$:

$$\boldsymbol{L} = \boldsymbol{D} + \boldsymbol{W}, \quad \boldsymbol{D} = \frac{1}{2}(\boldsymbol{L} + \boldsymbol{L}^T), \quad \boldsymbol{W} = \frac{1}{2}(\boldsymbol{L} - \boldsymbol{L}^T).$$

The physical interpretation of $\boldsymbol{D}$ and $\boldsymbol{W}$ is very simple. In terms of velocities, they correspond to the small strain tensor $\boldsymbol{\varepsilon}$ and small rotation tensor $\boldsymbol{\omega}$, respectively. Under the small displacement assumption, the kinematics can be considered as an incipient motion from the undeformed configuration.

The rate of the deformation gradient is computed as

$$\dot{\boldsymbol{F}} = \frac{d\boldsymbol{F}}{dt} = \frac{d}{dt}\left(\frac{d\boldsymbol{x}}{d\boldsymbol{X}}\right) = \frac{d}{d\boldsymbol{X}}\left(\frac{d\boldsymbol{x}}{dt}\right) = \frac{d\boldsymbol{v}}{d\boldsymbol{x}}\frac{d\boldsymbol{x}}{d\boldsymbol{X}} = \boldsymbol{L}\boldsymbol{F}. \tag{4.18}$$

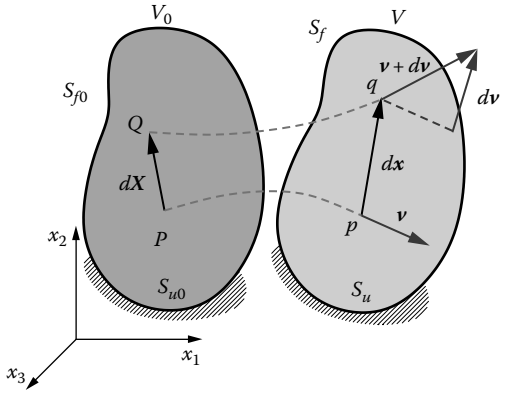


FIGURE 4.7 Meaning of the quantities $d\boldsymbol{v}$ and $d\boldsymbol{x}$ appearing in the velocity gradient $\boldsymbol{L}$.

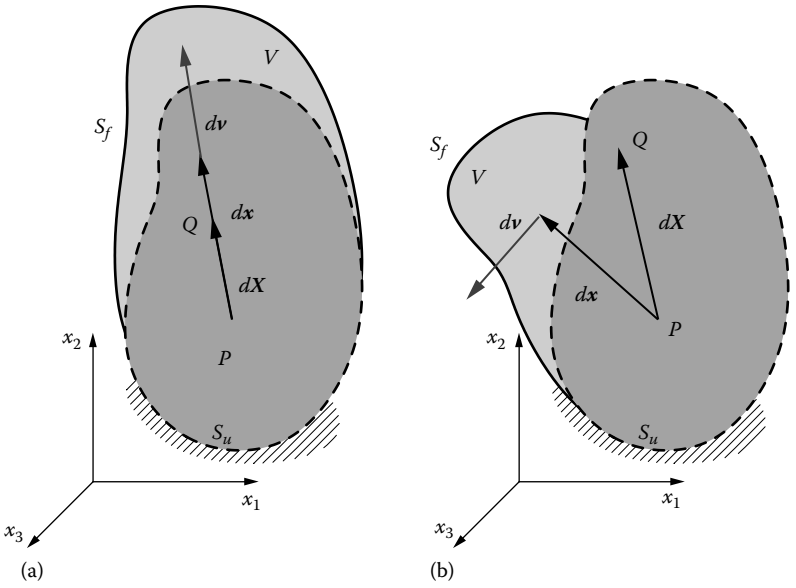


FIGURE 4.8 Examples of velocity vectors, leading to (a) deformation rate and (b) spin.

By using $(4.14)_1$ through (4.17) and (4.18), $\boldsymbol{D}$ and $\boldsymbol{W}$ are expressed as functions of the stretch tensor $\boldsymbol{U}$:

$$\boldsymbol{D} = \frac{1}{2}\boldsymbol{R}(\dot{\boldsymbol{U}}\boldsymbol{U}^{-1} + \boldsymbol{U}^{-1}\dot{\boldsymbol{U}})\boldsymbol{R}^T,$$

$$\boldsymbol{W} = \boldsymbol{\Omega}_{\boldsymbol{R}} + \frac{1}{2}\boldsymbol{R}(\dot{\boldsymbol{U}}\boldsymbol{U}^{-1} - \boldsymbol{U}^{-1}\dot{\boldsymbol{U}})\boldsymbol{R}^T$$

and, by using Equation 4.15, become

$$\begin{aligned} &\boldsymbol{D} = \boldsymbol{R}_E\boldsymbol{D}_E\boldsymbol{R}_E^T, \quad \boldsymbol{D}_E = \dot{\Lambda}\Lambda^{-1} + \frac{1}{2}\left(\Lambda^{-1}\boldsymbol{\Omega}_{\boldsymbol{L}}\Lambda - \Lambda\boldsymbol{\Omega}_{\boldsymbol{L}}\Lambda^{-1}\right), \\ &\boldsymbol{W} = \boldsymbol{R}_E\boldsymbol{W}_E\boldsymbol{R}_E^T, \quad \boldsymbol{W}_E = \boldsymbol{\Omega}_{\boldsymbol{E}} - \frac{1}{2}\left(\Lambda^{-1}\boldsymbol{\Omega}_{\boldsymbol{L}}\Lambda + \Lambda\boldsymbol{\Omega}_{\boldsymbol{L}}\Lambda^{-1}\right). \end{aligned} \tag{4.19}$$

$\boldsymbol{D}_E$ and $\boldsymbol{W}_E$ are the deformation rate and the spin tensors referred to the principal deformation axes at time t.

The diagonal elements of $(4.19)_2$ include only the first contribution of the left-side member. The second contribution is the difference between two transpose matrices and its diagonal terms are null. Therefore, the elements of $\dot{\boldsymbol{\Lambda}}$ are

$$[\dot{\Lambda}]_{\alpha\alpha} = \lambda_\alpha[\boldsymbol{D}_E]_{\alpha\alpha}.$$

For $\lambda_\alpha \neq \lambda_\beta$, the extra-diagonal elements of $\boldsymbol{\Omega}_L$ and $\boldsymbol{\Omega}_E$ are

$$[\boldsymbol{\Omega}_L]_{\alpha\beta} = \frac{2\lambda_\alpha\lambda_\beta}{\lambda_\beta^2 - \lambda_\alpha^2}[\boldsymbol{D}_E]_{\alpha\beta},$$

$$[\boldsymbol{\Omega}_E]_{\alpha\beta} = [\boldsymbol{W}_E]_{\alpha\beta} - \frac{\lambda_\alpha^2 + \lambda_\beta^2}{\lambda_\beta^2 - \lambda_\alpha^2}[\boldsymbol{D}_E]_{\alpha\beta}.$$

If $\boldsymbol{D}$ and $\boldsymbol{W}$ are described in the basis of the principal deformation system in the deformed configuration, where $\boldsymbol{R}_E = \boldsymbol{I}$, they define directly $\boldsymbol{\Omega}_L$ and $\boldsymbol{\Omega}_E$. Thus, $\dot{\boldsymbol{\Lambda}}$ is derived directly.

4.4.5 Strain Measures

Stretches measure the magnitude of deformations. They express the ratio between the length of a deformed segment and the length of the same segment undeformed. Thus, in the absence of deformation, they are equal to the unit. For small deformations, the stretch values differ slightly from the unit, say at the fourth or fifth digit. In engineering applications, it is more significant to adopt strain measures that are null in the case of null deformations, and assume nonzero values only when a deformation is present.

Different strain measures are available in finite kinematics. A general expression includes the definition of several strain tensors:

$$\boldsymbol{E}_G = \boldsymbol{R}_L g(\boldsymbol{\Lambda}) \boldsymbol{R}_L^T, \quad g(\boldsymbol{\Lambda}) = \mathrm{diag}[g(\lambda_i)],$$

where λ_i denotes the principal stretches. The corresponding strain rate tensor $\dot{\boldsymbol{E}}_G$ is given by

$$\dot{\boldsymbol{E}}_G = \boldsymbol{R}_L \dot{\boldsymbol{E}}_L \boldsymbol{R}_L^T, \quad \dot{\boldsymbol{E}}_L = \dot{\boldsymbol{\Lambda}} g'(\boldsymbol{\Lambda}) + \boldsymbol{\Omega}_L g(\boldsymbol{\Lambda}) - g(\boldsymbol{\Lambda}) \boldsymbol{\Omega}_L. \tag{4.20}$$

The earlier definition may be expanded in order to relate the components of $\dot{\boldsymbol{E}}_G$ with the components of $\boldsymbol{D}_E$, in the general form:

$$[\dot{\boldsymbol{E}}_G]_{\alpha\beta} = \gamma_{\alpha\beta} [\boldsymbol{D}_E]_{\alpha\beta}. \tag{4.21}$$

The explicit expression of $\gamma_{\alpha\beta}$ can be obtained only once $g(\boldsymbol{\Lambda})$ has been defined.

The most used strain measures in engineering applications are

1. Small strain tensor $\boldsymbol{\varepsilon}$:

$$\boldsymbol{\varepsilon} = \boldsymbol{R}_L (\boldsymbol{\Lambda} - \boldsymbol{I}) \boldsymbol{R}_L^T.$$

In Cartesian components, and introducing explicitly the displacements

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}).$$

2. Green–Lagrange strain tensor $\boldsymbol{E}$:

$$\boldsymbol{E} = \boldsymbol{R}_L \frac{1}{2}(\boldsymbol{\Lambda}^2 - \boldsymbol{I}) \boldsymbol{R}_L^T = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I}).$$

In Cartesian components, and introducing explicitly the displacements

$$E_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i} + u_{k,i} u_{k,j}).$$

The components of $\boldsymbol{E}$ do not change if the body undergoes a rigid rotation $\boldsymbol{R}$. In fact, the rigid rotation part $\boldsymbol{R}$ of $\boldsymbol{F}$ does not enter in the definition of $\boldsymbol{E}$.

The expression of $\gamma_{\alpha\beta}$ is

$$\gamma_{\alpha\beta} = \lambda_\alpha \lambda_\beta. \tag{4.22}$$

By combining Equations 4.19_1 through 4.22, the following relation is obtained:

$$\dot{\boldsymbol{E}} = \boldsymbol{F}^T \boldsymbol{D} \boldsymbol{F}, \quad \boldsymbol{D} = \boldsymbol{F}^{-T} \dot{\boldsymbol{E}} \boldsymbol{F}^{-1}, \tag{4.23}$$

also obtained by differentiating $\boldsymbol{E}$ with respect to time.

3. Hencky or logarithmic strain tensor ϵ:

$$\boldsymbol{\epsilon} = \boldsymbol{R}_L \log \boldsymbol{\Lambda} \boldsymbol{R}_L^T = \frac{1}{2} \log \boldsymbol{C}.$$

The expression of $\gamma_{\alpha\beta}$ is

$$\gamma_{\alpha\beta} = \begin{cases} 1 & \text{if } \lambda_\alpha = \lambda_\beta \\ \dfrac{2\lambda_\alpha \lambda_\beta}{\lambda_\alpha^2 - \lambda_\beta^2} \log \dfrac{\lambda_\beta}{\lambda_\alpha} & \text{otherwise} \end{cases}. \tag{4.24}$$

Because of the complexity of 4.24, it is not possible to find an explicit relation between ϵ and $\boldsymbol{D}$. The Hencky strain is used often in inelastic analyses. In that case, the correct incremental relation for ϵ is defined in terms of conjugate work.

Figure 4.9 compares the three strain measures ε_{ij}, ϵ_{ij}, and E_{ij} in the uniaxial case. In small strain condition (i.e., when the stretch is close to one), the three measures coincide.

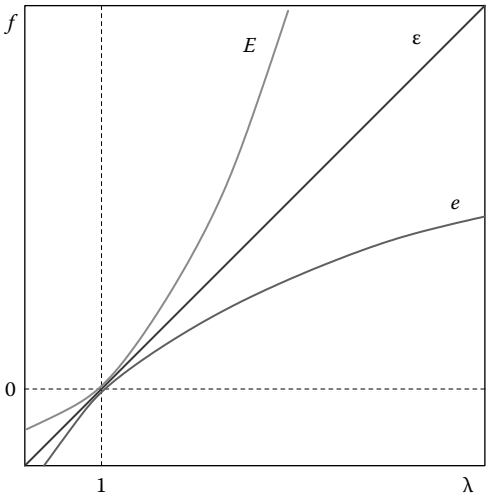


FIGURE 4.9 Comparison of the deformation measures in the uniaxial case. The lowest curve corresponds to the logarithmic deformation. The straight curve corresponds to the engineering deformation (small strain). The upper curve corresponds to Green–Lagrange.

An example of Eulerian strain measure is the Almansi strain tensor. It relates with the Green–Lagrange strain tensor as follows:

$$\boldsymbol{A} = \frac{1}{2}[\boldsymbol{I} - \boldsymbol{B}^{-1}] = \frac{1}{2}[\boldsymbol{I} - \boldsymbol{F}^{-T} \cdot \boldsymbol{F}^{-1}] = \boldsymbol{F}^{-T} \cdot \boldsymbol{E} \cdot \boldsymbol{F}^{-1}.$$

In Cartesian components, it reads

$$A_{ij} = \frac{1}{2}\left[u_{i,j} + u_{j,i} - u_{k,i}u_{k,j}\right].$$

Unlike Green–Lagrange strain tensor, the Almansi strain tensor components are not invariant under rigid rotations.

4.4.6 Stress Measures

The stress measures for finite deformations are dependent on the choice of the strain measures. The Cauchy stress $\boldsymbol{\sigma}$ is defined in the current configuration and is suitable for Eulerian formulations. In a Lagrangian formulation, the most used stress measures are the first $\boldsymbol{P}$ and the second $\boldsymbol{S}$ Piola–Kirchhoff tensors. The two Piola–Kirchhoff tensors do not have a precise physical meaning. They are introduced in order to provide a traction, that is, a force per unit of undeformed surface. Therefore, the only physically meaningful stress measure also in large strains is the Cauchy stress.

Figure 4.10 shows two traction vectors. The surface traction $d\boldsymbol{t}$ is applied to the elementary deformed surface dS, while the surface traction $d\boldsymbol{T}$ is applied to the elementary undeformed surface dS^0. The first Piola–Kirchhoff tensor $\boldsymbol{P}$ refers the current surface traction $d\boldsymbol{t}$ to the undeformed surface dS^0. Thus, the traction $d\boldsymbol{t}$ is linked unnaturally to the normal $\boldsymbol{N}$ to dS^0 in the undeformed reference system. By extending the definition of Cauchy stress to the undeformed surface, one obtains

$$d\boldsymbol{t} = (\boldsymbol{n}\boldsymbol{\sigma})dS = (\boldsymbol{N}\boldsymbol{P})dS^0.$$

The second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ uses the surface traction $d\boldsymbol{T}$, related to $d\boldsymbol{t}$ by the same relation that links $d\boldsymbol{X}$ and $d\boldsymbol{x}$, that is, the inverse of the deformation gradient $\boldsymbol{F}^{-1}$. It also refers it to the undeformed surface dS^0:

$$d\boldsymbol{T} = \boldsymbol{F}^{-1}d\boldsymbol{t} = \boldsymbol{F}^{-1}(\boldsymbol{n}\boldsymbol{\sigma})dS = (\boldsymbol{N}\boldsymbol{S})dS^0.$$

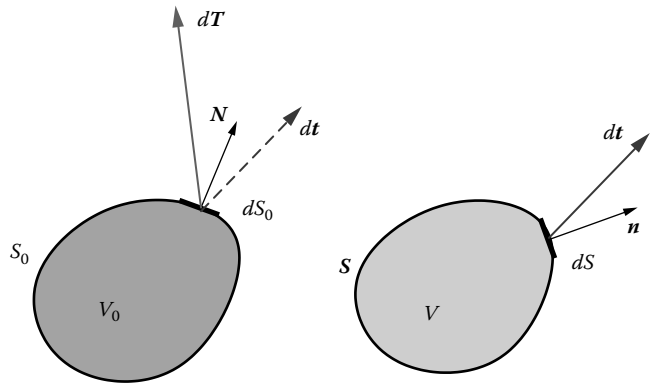


FIGURE 4.10 Traction vectors for the definition of the Piola–Kirchhoff stress tensors.

To find the relations between $\boldsymbol{\sigma}$, $\boldsymbol{P}$, and $\boldsymbol{S}$, it is necessary to express the deformed area dS as a function of the undeformed one dS^0 through the Nanson's formula [2]:

$$\boldsymbol{n}\, dS = \boldsymbol{JNF}^{-1}\, dS^0,$$

obtaining

$$\boldsymbol{\sigma} = J^{-1}\boldsymbol{FSF}^T = J^{-1}\boldsymbol{FP}. \tag{4.25}$$

Unlike $\boldsymbol{\sigma}$ and $\boldsymbol{S}$, $\boldsymbol{P}$ is nonsymmetric.

A strain tensor and a stress tensor are said to be work conjugated if the sum of the product of their equal-index components is a physical work or energy per unit of volume. In a constitutive relationship, only work-conjugated tensors can be used. The second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ is work conjugated to the Green–Lagrange strain tensor $\boldsymbol{E}$, while the first Piola–Kirchhoff stress tensor $\boldsymbol{P}$ is work conjugated to the deformation gradient $\boldsymbol{F}$.

$\boldsymbol{S}$ and $\boldsymbol{E}$ are the most widely used pair of tensors in finite deformation analyses. Both $\boldsymbol{S}$ and $\boldsymbol{E}$ do not change under rigid rotations, since a rigid rotation does not affect the Cauchy–Green strain tensor $\boldsymbol{C}$. As a consequence of their definition, the components of $\boldsymbol{S}$ and $\boldsymbol{E}$ do not change if a rigid rotation is applied [1]. Such property, called material frame indifference, does not hold for the stress rates.

4.5 NONLINEAR FORMULATIONS OF THE VIRTUAL WORK EQUATION

For a generic time t, $0 \le t \le T$, of a loading process, the internal virtual work in Equation 4.9_1 can be written in terms of 4.25 and 4.23, accounting for $\rho dV = \rho^0 dV^0$:

$$\int_V \sigma_{ij}\,\delta e_{ij}\, dV = \int_V \left(\frac{\rho}{\rho^0} F_{ih} S_{hk} F_{jk}\right)\left(F_{il}^{-1}\delta E_{lm} F_{jm}^{-1}\right) dV = \int_{V^0} S_{hk}\delta E_{hk}\, dV^0. \tag{4.26}$$

A similar treatment is needed for the external work W, whose expression has to be derived according to the particular case in examination.

Closed form solutions of 4.9 or 4.26, referred to the unknown configuration at time t, cannot be derived. Solutions are obtained in approximate form. In general, the solution is evaluated at predefined times, separated by constant intervals Δt (e.g., $t = t_0 + \Delta t$). The approximation is based on the linearization of the equations, and needs to know the solutions at previous times. Then, the accuracy is improved by iterative procedures. In practice, two formulations have become more popular than others, that is, the TL formulation and the UL formulation:

- The TL formulation refers the static and kinematic quantities to the initial configuration at time $t = 0$:

$$\int_{V^0} S_{hk}(t)\,\delta E_{hk}(t)\, dV^0 = W^0(t). \tag{4.27}$$

- The UL formulation refers the static and kinematic quantities to the configuration at time $t = t_0$, and the deformation is expressed as increment $\delta\Delta E_{hk}$ between t_0 and t:

$$\int_{V^t} S_{hk}(t)\delta\Delta E_{hk}\, dV^t = W^t(t).$$

Both formulations include all the effects of finite kinematics, that is, large displacements, large rotations, and large deformations, but the correctness of the approach depends on the adopted constitutive relations. The choice of the formulation is motivated by the computational efficiency.

The formulations are based on the consistent linearization of governing equations, stresses, and strains.

4.5.1 TL Formulation

The TL formulation of the virtual work equation is based on the following steps:

1. The stress $S_{hk}(t)$ is expressed in terms of the initial configuration variables at time $t = 0$:

$$S_{hk}(t) = \frac{\rho^0}{\rho} F_{ih}^{-1} \sigma_{ij}(t) F_{jk}^{-1}, \quad F_{ih}^{-1} = \frac{dX_i}{dx_h(t)}.$$

2. Also, the deformation $E_{hk}(t)$ is expressed as a function of the initial configuration variables at time $t = 0$:

$$E_{hk}(t) = \frac{1}{2}(u_{h,k}(t) + u_{k,h}(t) + u_{j,h}(t)u_{j,k}(t)), \quad u_h(t) = x_h(t) - X_h.$$

3. The stress is written as the sum of the value at the previous step and of an increment ΔS_{hk} referred to the initial configuration:

$$S_{hk}(t) = S_{hk}(t_0) + \Delta S_{hk}.$$

4. The displacement is written as the sum of the value at the previous step and of an increment Δu_h referred to the initial configuration:

$$u_h(t) = u_h(t_0) + \Delta u_h. \tag{4.28}$$

The consistent deformation is written as the sum of three contributions:

$$E_{hk}(t) = E_{hk}(t_0) + \Delta E_{hk} = E_{hk}(t_0) + \Delta e_{hk} + \Delta \eta_{hk}. \tag{4.29}$$

a. A known term (with null variation):

$$E_{hk}(t_0) = \frac{1}{2}(u_{h,k}(t_0) + u_{k,h}(t_0) + u_{j,h}(t_0)u_{j,k}(t_0)).$$

b. A term linear in the increment $\Delta u_{h,k}$ (accounts for the initial displacement):

$$\Delta e_{hk} = \frac{1}{2}(\Delta u_{h,k} + \Delta u_{k,h} + \Delta u_{j,h}u_{j,k}(t_0) + u_{j,h}(t_0)\Delta u_{j,k}).$$

c. A term quadratic in the increment $\Delta u_{h,k}$:

$$\Delta \eta_{hk} = \frac{1}{2}\Delta u_{j,h}\,\Delta u_{j,k}.$$

5. The variation of the deformation (4.29) is inserted in the virtual work equation:

$$\delta E_{hk} = \delta \Delta E_{hk} = \delta \Delta e_{hk} + \delta \Delta \eta_{hk},$$

Obtaining

$$\int_{V^0} \Delta S_{hk} \delta \Delta E_{hk} \, dV^0 + \int_{V^0} S_{hk}(t_0) \delta \Delta \eta_{hk} \, dV^0 = W^0(t) - \int_{V^0} S_{hk}(t_0) \delta \Delta e_{hk} \, dV^0. \tag{4.30}$$

The left side of 4.30 is strongly nonlinear in the displacement increments. The right side of 4.30 contains all the known terms, the integral being known when the current displacement is known.

6. The second integral in the right side is a linear function of $\Delta u_{h,k}$ and does not need changes. The first integral instead is linearized by Taylor's expansion of the stress S_{hk} and by using 4.29:

$$\Delta S_{hk} \approx \left. \frac{\partial S_{hk}}{\partial E_{ij}} \right|_t \Delta E_{ij} = C^t_{hkij} (\Delta e_{ij} + \Delta \eta_{ij}).$$

The material tangent stiffness C^t_{hkij} at time t is referred to the configuration at time t_0. Finally, the increment of the deformation is linearized:

$$\Delta E_{ij} \approx \Delta e_{ij}.$$

At this point, the first integral in 4.30 is a linear function of the displacement increment $\Delta u_{h,k}$. This gives the virtual work equation in the TL formulation:

$$\int_{V^0} C^t_{hkij} \Delta e_{ij} \delta \Delta e_{hk} \, dV^0 + \int_{V^0} S_{hk}(t_0) \delta \Delta \eta_{hk} \, dV^0 = W^0(t) - \int_{V^0} S_{hk}(t_0) \delta \Delta e_{hk} \, dV^0. \tag{4.31}$$

4.5.2 UL Formulation

The UL formulation of the virtual work theorem is based on the following steps:

1. The stress $S_{hk}(t)$ is expressed as function of the configuration variables at the reference time t_0:

$$S_{hk}(t) = \frac{\rho^{t_0}}{\rho} F^{-1}_{ih} \sigma_{ij}(t) F^{-1}_{jk}, \quad F^{-1}_{ih} = \frac{dx_i(t_0)}{dx_h(t)}.$$

2. Since the deformation is referred to the previous step, its value coincides with the increment of the deformation ΔE_{hk}. The deformation is expressed as a function of the displacement $u_h(t) = \Delta u_h$ referred to the previous step:

$$\Delta E_{hk} = \frac{1}{2}(\Delta u_{h,k} + \Delta u_{k,h} + \Delta u_{j,h}\Delta u_{j,k}), \quad \Delta u_h = u_h(t) - u_h(t_0).$$

3. The stress is decomposed into the sum of two terms. The first is the stress at the end of the previous step, referred to the previous step itself. It corresponds to the Cauchy stress σ_{hk} at the previous time. The second term is the increment from the previous time step:

$$S_{hk}(t) = \sigma_{hk}(t_0) + \Delta S_{hk}.$$

4. Following the same decomposition as in TL, the strain is expressed now as the sum of two contributions:

$$\Delta E_{hk} = \Delta e_{hk} + \Delta \eta_{hk}. \tag{4.32}$$

   a. A term linear in the displacement increment $\Delta u_{h,k}$:

$$\Delta e_{hk} = \frac{1}{2}(\Delta u_{h,k} + \Delta u_{k,h}).$$

   b. A term quadratic in the displacement increment $\Delta u_{j,h}$:

$$\Delta \eta_{hk} = \frac{1}{2}\Delta u_{j,h}\,\Delta u_{j,k}.$$

5. The virtual work equation is written in terms of variation of the incremental strain:

$$\int_{V^t} \Delta S_{hk}\delta\Delta E_{hk}\, dV^t + \int_{V^t} \sigma_{hk}(t_0)\,\delta\Delta\eta_{hk}\, dV^t = W^t(t) - \int_{V^t} \sigma_{hk}(t_0)\,\delta\Delta e_{hk}\, dV^t. \tag{4.33}$$

   The left side of 4.33 is strongly nonlinear in the displacement increment. The right-hand side contains all the known terms, since the integral is known when the displacement is known.

6. By linearizing the constitutive relations, the incremental stress is written as $\Delta S_{hk} = C^t_{hkij}\Delta e_{ij}$ (where the material tangent stiffness C^t_{hkij} at time t is referred to the configuration at time t) and the strains are written as $\delta\Delta E_{hk} = \delta\Delta e_{hk}$. The virtual work equation in the UL formulation is

$$\int_{V^t} C^t_{hkij}\Delta e_{ij}\,\delta\Delta e_{hk}\, dV^t + \int_{V^t} \sigma_{hk}(t_0)\,\delta\Delta\eta_{hk}\, dV^t = W^t(t) - \int_{V^t} \sigma_{hk}(t_0)\,\delta\Delta e_{hk}\, dV^t. \tag{4.34}$$

The formulations 4.31 and 4.34 are very similar. The only difference is conceptual and is the choice of the reference system, where the static and kinematics variables are expressed. The results obtained from the two formulations have to be identical, if the correct stress and strain measures are used. The choice of TL or UL is matter of computational efficiency. In general, the TL formulation provides a more complex stiffness matrix, due to the presence of an initial displacement.

Note that in the case that the nonlinearities can be confined to the material (i.e., small displacements), a less demanding approach can be followed. The virtual work equation can be written in the undeformed configuration (neglecting the geometry changes) and the strain accounts only for the linear components. In that case, TL and UL reduce to the same expression:

$$\int_V C_{hkij}^{t(k-1)} \Delta e_{ij}^{(k)} \delta \Delta e_{hk} \, dV = W(t) - \int_V \sigma_{hk}^{(k-1)} \delta \Delta e_{hk} \, dV.$$

In particular, for a linear elastic material it coincides with the classic virtual work theorem.

In conclusion, Equations 4.31 and 4.34 are used for the computation of the incremental displacements at time t. The total displacements are approximated through 4.28; strains are evaluated through 4.29 or 4.32, and stresses through the constitutive equations, described in the next section.

4.6 CONSTITUTIVE LAWS

For finite displacement formulations, it is necessary to use the proper constitutive relationships. If the finite deformation behavior is described through a TL approach, the constitutive law has to be expressed in terms of finite strains.

Table 4.2 summarizes the different models of behavior to be used in finite deformations. The internal variables q_h are scalar, vector, or tensor variables that have to be introduced in the constitutive behavior of nonlinear materials to keep track of the loading history.

The nonlinear analysis is conducted in incremental form, and direct approaches are excluded. At each time increment Δt, a nonlinear analysis requires the iteration of a sequence of steps, until the convergence test is satisfied. Clearly, the solution at the end of the previous step, that is, $S_{ij}(t_0)$, $E_{ij}(t_0)$, and $q_h(t_0)$, is known:

1. From the known nodal displacements $\boldsymbol{U}(t)^{(k-1)}$, the corresponding strains $\boldsymbol{E}(t)^{(k-1)}$ are computed.
2. Compute in sequence:

TABLE 4.2
Models of Material Behaviors

Material Model	Features	Materials
Linear elastic	$\sigma_{ij} = C_{ijhk}\varepsilon_{hk}$	Steel, $\sigma_{ij} < \sigma_y$
Nonlinear elastic	$\sigma_{ij} = C_{ijhk}(\varepsilon_{hk})\varepsilon_{hk}$	Aluminum
Hyper-elastic	$S_{ij} = \partial W/\partial E_{ij}$	Rubber
Hypo-elastic	$d\sigma_{ij} = C_{ijhk} d\epsilon_{hk}$	Concrete
Elastoplastic	$dS_{ij} = C_{ijhk}^{t}(E_{hk}, q_h) dE_{hk}$	Steel, $\sigma_{ij} \leq \sigma_y$
Creep	$S_{ij} = C_{ijhk}(E_{hk}, t)E_{hk}$	High-temperature metals
Visco-plastic	$dS_{ij} = C_{ijhk}(E_{hk}, \dot{E}_{hk}, q_h) dE_{hk}$	Polymers

- The stresses $\boldsymbol{S}(t)^{(k-1)}$. In elastic analysis, the stresses are computed directly from the strains $\boldsymbol{E}(t)^{(k-1)}$, at the end of the iteration $(k-1)$. In inelastic analysis, strains are computed from the integration of the equation:

$$\boldsymbol{S}(t)^{(k-1)} = \boldsymbol{S}(t_0) + \int_{t_0}^{t^{(k-1)}} d\boldsymbol{S}. \tag{4.35}$$

- The tangent stiffness of the material $\boldsymbol{C}^{(k-1)}$. In elastic analysis, it is obtained directly from the strains $\boldsymbol{E}(t)^{(k-1)}$, at the end of the iteration $(k-1)$. In inelastic analysis, the stiffness matrix refers to the state t, at the end of the iteration $(k-1)$. It has to be computed consistently to the integration process (4.35). If isoparametric finite elements are used, strains and stresses have to be computed in all the integration points of the mesh, in order to define the equations used in step 3.
- The internal variables $\boldsymbol{q}(t)^{(k-1)}$ (only for inelastic analyses).

3. The stresses and the material stiffness matrix are numerically integrated over each elements, building the local nodal force array $\boldsymbol{F}^{\text{int}}(t)^{(k-1)}$ and the local element stiffness matrix $\boldsymbol{K}(t)^{(k-1)}$.
4. The increments of the nodal displacements $\Delta\boldsymbol{U}^{(k)}$ at the (k) iteration are computed as

$$\boldsymbol{K}(t)^{(k-1)}\Delta\boldsymbol{U}^{k} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t)^{(k-1)}$$

$$\boldsymbol{U}(t)^{k} = \boldsymbol{U}(t)^{(k-1)} + \Delta\boldsymbol{U}^{k}.$$

Steps 1 to 4 are repeated to convergence.

The procedure works if all the variables (strains, stresses, and internal variables) at the end of the previous iteration are known. The variables have to be updated at all the integration points and at each iteration; therefore, it is incumbent to adopt the most efficient computational procedure.

The main difference between a nonlinear elastic analysis and an inelastic analysis is that in the former the total variables can be used, and an incremental formulation is not necessary.

4.6.1 Linear Elastic Behavior

Although not realistic, the simplest material model in finite deformations is the Hooke law. In a TL approach, it reads

$$S_{ij} = C_{ijhk}E_{hk}, \tag{4.36}$$

where, with respect to the linear approach, the Green–Lagrange strain tensor $\boldsymbol{E}$ replaces the small strain tensor $\boldsymbol{\varepsilon}$ and the second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ replaces the Cauchy stress tensor $\boldsymbol{\sigma}$. Formally, 4.36 coincides with 4.1 except for the stress and strain measures.

In a large displacements, large rotations, and small strain formulation, relation 4.36 is a natural material (Lagrangian) description. In fact, the components of the Green–Lagrange strain tensor $\boldsymbol{E}$ and the second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ measured in a fixed reference system do not change when the body undergoes rigid rotations and translations. Therefore, only stretches induce stress increments.

The earlier remark is important not only for the linear elastic behavior. In fact, any constitutive law developed in the material description for small displacement analysis (i.e., expressed in terms of small strain tensor $\boldsymbol{\varepsilon}$ and the Cauchy stress tensor $\boldsymbol{\sigma}$) can be used directly for large displacement,

large rotations, and small strains analyses, if the Green–Lagrange strain tensor $\boldsymbol{E}$ replaces $\boldsymbol{\varepsilon}$ and the second Piola–Kirchhoff stress tensor $\boldsymbol{S}$ replaces $\boldsymbol{\sigma}$.

Such consideration is of great practical interest, since many engineering problems imply large displacements and large rotations but can be treated in small strains. This is, for example, the case of buckling in slender beams and thin shells.

In order to use 4.36 in the UL formulation, the virtual work equation (4.27) has to be written as

$$\int_{V^0} C_{ijhk}(t_0)E_{hk}(t_0)\delta E_{ij}(t_0)\,dV^0 = W^0(t_0). \tag{4.37}$$

A new constitutive tensor, C_{mnrs}, referred to the current configuration, is then introduced:

$$C_{mnrs} = \frac{\rho(t_0)}{\rho^0} F_{mi}F_{nj}C_{ijhk}(t_0)F_{rh}F_{sk}, \quad C_{ijhk}(t_0) = \frac{\rho^0}{\rho(t_0)} F_{im}^{-1}F_{jn}^{-1}C_{mnrs}F_{hr}^{-1}F_{ks}^{-1}. \tag{4.38}$$

It is easy to prove that the variation of the small strain tensor in the current configuration is expressed as [1]:

$$\delta\varepsilon_{mn} = F_{im}^{-1}F_{jn}^{-1}\delta E_{ij}. \tag{4.39}$$

Replacing 4.38 and 4.39 in 4.37, and converting the integral over the initial volume into the integral over the current volume, one obtains

$$\int_{V^t} C_{mnrs}A_{rs}\delta\varepsilon_{mn}(t_0)\,dV^t = W(t_0),$$

where A_{rs} denotes the components of the Almansi strain tensor and

$$\sigma_{mn} = C_{mnrs}A_{rs} \tag{4.40}$$

is the Hooke law written in the UL approach. Practically, the same result can be obtained by transforming $\boldsymbol{S}$ in $\boldsymbol{\sigma}$, and using the relation (4.9). This avoids the calculation of C_{mnrs} and is computationally more efficient.

The equivalence between TL and UL formulations holds for any stress level, although the Hooke law is rigorously valid only in the small strain range.

It is important to observe that if the same constants are used to define 4.36 in TL and 4.40 in UL, no relevant differences in the results are observed for small strains. If the strains are large, instead, the two approaches give quite different results. But, if Equation 4.38 is used to switch from the material constants of one formulation to the material constants of the other formulation, the same response is obtained.

4.6.2 Nonlinear Elastic Behavior

Materials exhibiting large strains in their standard applications (see, e.g., rubber) and a nonlinear reversible behavior (hyper-elastic) can be characterized by a strain energy density function W such that

$$S_{ij} = \frac{\partial W}{\partial E_{ij}}, \quad C_{ijhk} = \frac{\partial^2 W}{\partial E_{ij}\partial E_{hk}}.$$

If the material is isotropic, W can be expressed as a function of the eigenvalues of $\boldsymbol{C}$, λ_i^2, or, alternatively, of the $\boldsymbol{C}$'s invariants I_1, I_2, and I_3 [5,6]:

$$I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2, \quad I_2 = \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2, \quad I_3 = \lambda_1^2\lambda_2^2\lambda_3^2 = J^2.$$

Thus, an isotropic material strain energy function can be written as

$$W = W(\lambda_1^2, \lambda_2^2, \lambda_3^2) = W(I_1, I_2, I_3).$$

Assuming that W is continuum, smooth, and differentiable, it can be expressed through a power series:

$$W(I_1, I_2, I_3) = \sum_{p,q,r=0}^{\infty} c_{pqr}(I_1 - 3)^p (I_2 - 3)^q (I_3 - 1)^r, \tag{4.41}$$

where the coefficients c_{pqr} are independent of the strains. In particular, the condition $W = 0$ in the undeformed configuration implies that $c_{000} = 0$ and $c_{100} + 2c_{010} + c_{001} = 0$.

A special class of materials is obtained by excluding the dependence on the invariant I_2:

$$W(I_1, I_3) = c_1(I_1 - 3) + g(I_3).$$

The class of materials is known as neo-Hookean extended to the compressible range. An example of strain energy density with a particularly satisfying choice of $g(I_3)$ is

$$W(I_1) = \frac{1}{2}\mu(I_1 - 3 - 2\log J) + \frac{1}{2}K(J-1)^2,$$

where K is a coefficient related to the bulk modulus of the material.

4.6.2.1 Incompressible Materials

Rubberlike materials exhibit a quasi-incompressible behavior [5,6]. The incompressibility constraint (no volume change) is expressed by the condition:

$$I_3 = \lambda_1^2\lambda_2^2\lambda_3^2 = 1, \quad \lambda_3^2 = \lambda_1^{-2}\lambda_2^{-2}.$$

For such materials, Equation 4.41 can be written as

$$W(I_1, I_2) = \sum_{p,q=0}^{\infty} c_{pq}(I_1 - 3)^p (I_2 - 3)^q.$$

Among the most diffused material models for rubber, it is worth to mention the following:

- Mooney–Rivlin model:

$$W(I_1, I_2) = \frac{1}{2}\mu_1(I_1 - 3) - \frac{1}{2}\mu_2(I_2 - 3), \quad \mu = \mu_1 - \mu_2,$$

also written as

$$W(\lambda_1^2, \lambda_2^2) = \frac{1}{2}\mu_1(\lambda_1^2 + \lambda_2^2 + \lambda_1^{-2}\lambda_2^{-2} - 3) - \frac{1}{2}\mu_2(\lambda_1^{-2} + \lambda_2^{-2} + \lambda_1^2\lambda_2^2 - 3),$$

where μ is the shear elastic modulus G and $E = 6(\mu_1 + \mu_2)$.

- Neo-Hookean model:

$$W(I_1) = \frac{1}{2}\mu(I_1 - 3),$$

also written as

$$W(\lambda_1^2, \lambda_2^2) = \frac{1}{2}\mu(\lambda_1^2 + \lambda_2^2 + \lambda_1^{-2}\lambda_2^{-2} - 3).$$

- Odgen model:

$$W(\lambda_1, \lambda_2) = \sum_{n=1}^{3} \frac{\mu_n}{\alpha_n}(\lambda_1^{\alpha_n} + \lambda_2^{\alpha_n} + \lambda_1^{-\alpha_n}\lambda_2^{-\alpha_n} - 3),$$

where μ_n and α_n are material constants, and the following relation holds

$$2\mu = \sum_{n=1}^{3} \alpha_n \mu_n.$$

A comparison between the material response and an uniaxial loading for different rubber material model is reported in Figure 4.11.

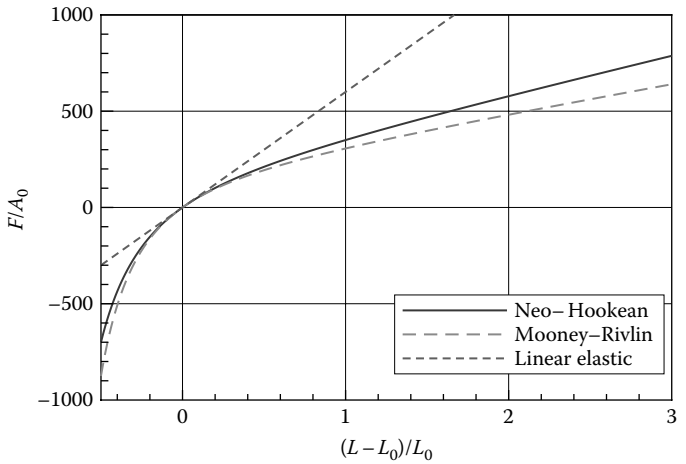


FIGURE 4.11 Uniaxial response of a rubber, compared to the linear response of the Hooke law.

The use of incompressible material models introduces numerical complications. Such phenomena are known as volumetric locking. To avoid the difficulties linked to such formulations, mixed approaches are used. Mixed approaches discretize as primary variables both the displacement field and a static field, for example, the isotropic pressure [1].

4.6.3 Inelastic Behavior

A material behavior is said inelastic when the value of the stress does not depend only on the strain but also on the loading history. Examples of inelastic behaviors are elastoplasticity, creep, viscoplasticity. The dependence on the loading history requires the adoption of an incremental analysis able to follow the loading history.

As already mentioned, there are three possible kinematic formulations:

- SD-SR-SS, small displacements and rotations, and small strains. A MN formulation is used, where the nonlinearities are related only to the material behavior.
- LD-LR-SS, large displacements and rotations, and small strains. The TL formulation, which can be used also for large strain problems, is employed successfully. If small strains are assumed, the same material model developed for MN formulations can be used, by simply replacing the strain and stress engineering measures with Green–Lagrange and second Piola–Kirchhoff tensors, respectively.
- LD-LR-LS, large displacements and large deformations. The TL formulation or the UL formulation can be used without substantial difference. It is important to observe that in this case the constitutive equations are more complicated, and numerous operative difficulties are encountered.

Since the first two kinematic formulations adopt the same constitutive relationships, in the following, only the MN formulation is discussed. The large strain for the TL formulations is discussed next.

The complete and accurate solution, in terms of displacements, strains, stresses, and internal variables, has been obtained at time t. All the iterative schemes are based on two subsequent steps:

1. Computation of the tangent stiffness of the material at time t, written as

$$C^t_{ijhk} = \left.\frac{\partial \sigma_{ij}}{\partial \varepsilon_{hk}}\right|_t . \tag{4.42}$$

   The tangent stiffness of the material is used in the evaluation of the global tangent stiffness matrix of the system. If the calculation of the tangent stiffness is not accurate, the displacement increments obtained in the iteration are not accurate, and the convergence rate of the algorithm is reduced.
2. Computation of stresses, inelastic strains, internal variables at time t. If the calculation of stresses is not accurate, the numeric error, in general, cannot be recovered in the next iterations. In practice, any error introduced in the evaluation of $\sigma_{ij}(t)$, or inelastic strains, or internal variables, contributes to spoil the accuracy of the method. Therefore, it compromises the ability of the algorithm to predict the mechanical response of the system.

   During the equilibrium iteration, the tangent stiffness and the stresses have to be computed in a consistent way at the current stress state. Consistency, formally expressed by Equation 4.42, defines the tangent stiffness as the partial derivative of the stresses with respect to the strains. Partial differentiation requires to consider all the dependencies, explicit and implicit, on the strains. The consistency concept is very important for the efficiency of the numerical approaches, and it will be recalled later.

4.6.3.1 Small Strain Elastoplasticity (von Mises)

Elastoplasticity theories in small strains, for example, Prandtl–Reuss equations [2,7], are based on the assumption of the additive decomposition of the incremental strains into en elastic $d\varepsilon_{ij}^{e}$ part and a plastic $d\varepsilon_{ij}^{p}$ part:

$$d\varepsilon_{ij} = d\varepsilon_{ij}^{e} + d\varepsilon_{ij}^{p}.$$

Only the elastic part of the strains contributes to the stresses through the elasticity tensor C_{ijhk}^{e}:

$$d\sigma_{ij} = C_{ijhk}^{e}(d\varepsilon_{ij} - d\varepsilon_{ij}^{p}). \tag{4.43}$$

The incremental total strains in general are known data of the problem. To compute the irreversible or inelastic part of the strains, the material model has to be completed with the inelastic constitutive relations:

1. A plasticity stress-dependent function, to define the combination of multiaxial stress states leading to the development of plastic strains. The plasticity function is written as

$$f_y(\sigma_{ij}, \varepsilon_{ij}^{p}, q_\alpha) \le 0.$$

   The material response is assumed to be elastic if $f_y < 0$. The response is either elastic or plastic, according to the sign of the incremental strains, if $f_y = 0$. The latter condition is called plasticity condition, and identifies the states where plastic strains can develop.
2. A flow rule, to define the increment of the plastic strains, function of the stresses and of the stress increments. The flow rule describes the evolution of the plastic strains. The material may exhibit an associate plastic behavior. In this case, the plastic strains are directed as the gradient of f_y, that is, they are parallel to the normal to the plasticity function. An associate plastic flow rule reads

$$d\varepsilon_{ij}^{p} = \frac{\partial f_y(\sigma_{ij}, \varepsilon_{ij}^{p})}{\partial \sigma_{ij}}\lambda,$$

   where λ is a scalar that measures the plastic strain magnitude.
3. A hardening or softening rule, to describe the evolution of the plasticity function with the plastic strains or with more general inelastic phenomena (e.g., fracture, damage, etc.). The hardening/softening rule modifies the value of the internal variables q_α used in f_y and the plasticity condition.

   In the standard von Mises plasticity, the volumetric plastic deformations are zero. In order to exploit this feature, it is convenient to split the elastic constitutive relation (4.43) into deviatoric and volumetric parts:

$$\boldsymbol{s}(t) = 2\mu[\boldsymbol{e}(t) - \boldsymbol{e}^{p}(t)], \quad \mu = \frac{E}{2(1+\nu)}, \tag{4.44}$$

$$\sigma_m(t) = 3K\varepsilon_m(t), \quad K = \frac{E}{3(1-2\nu)}, \tag{4.45}$$

   where
   $\boldsymbol{s}$ is the deviatoric stress
   σ_m is the mean stress (or isotropic pressure):

$$s_{ij} = \sigma_{ij} - \sigma_m \delta_{ij}, \quad \sigma_m = \frac{\sigma_{kk}}{3},$$

$\boldsymbol{e}$ is the deviatoric strain
ε_m is one third of the volumetric deformation:

$$e_{ij} = \varepsilon_{ij} - \varepsilon_m \delta_{ij}, \quad \varepsilon_m = \frac{\varepsilon_{kk}}{3}.$$

The deviatoric stress (4.44) is a function of the plastic strains, which in turn are dependent on the loading history. The mean stress is purely elastic. The only unknowns of the problem are the plastic strains $\boldsymbol{e}^p$ and the deviatoric stresses $\boldsymbol{s}$.

The total plastic strains can be written as the sum of a known term $\boldsymbol{e}^p(t_0)$ and of an unknown incremental term $\Delta\boldsymbol{e}^p$:

$$\boldsymbol{e}^p(t) = \boldsymbol{e}^p(t_0) + \Delta\boldsymbol{e}^p.$$

Thus, the elastic part of the deviatoric relationship is written as

$$\boldsymbol{s}(t) = 2\mu[\boldsymbol{e}''(t) - \Delta\boldsymbol{e}^p] \quad \boldsymbol{e}'' = \boldsymbol{e}(t) - \boldsymbol{e}^p(t_0). \tag{4.46}$$

Both $\Delta\boldsymbol{e}^p$ and $\boldsymbol{s}$ have to satisfy the plasticity condition, the flow rule, and the hardening rule. The von Mises plasticity function reads

$$f_y^M(t) = \frac{1}{2}\boldsymbol{s}(t)\cdot\boldsymbol{s}(t) - \frac{1}{3}\sigma_y^2(t), \tag{4.47}$$

where
$\boldsymbol{s}(t)\cdot\boldsymbol{s}(t) = s_{ij}s_{ij}$
$\sigma_y(t)$ is the current yield stress, function of the hardening of the material:

$$\sigma_y(t) = g(\bar{e}^p), \quad \bar{e}^p = \int_0^t \sqrt{\frac{2}{3} d\boldsymbol{e}^p \cdot d\boldsymbol{e}^p}\, dt.$$

During the evolution of the plastic strains, the stresses have to satisfy the plasticity condition $f_y^M = 0$ (see Figure 4.12). The associated flow rule states that the increment of the plastic strains has to be parallel to the final deviatoric stresses $\boldsymbol{s}(t)$, while the volumetric plastic strains are zero:

$$\Delta\boldsymbol{e}^p = \lambda \frac{\partial f_y^M(t)}{\partial\sigma} = \lambda\boldsymbol{s}(t). \tag{4.48}$$

The plastic multiplier λ is obtained by taking the tensor product of both the sides of 4.48:

$$\Delta\bar{e}^p = \frac{2}{3}\lambda\bar{\sigma}(t), \quad \Delta\bar{e}^p = \sqrt{\frac{2}{3}\Delta\boldsymbol{e}^p\cdot\Delta\boldsymbol{e}^p}, \quad \bar{\sigma}(t) = \sqrt{\frac{3}{2}\boldsymbol{s}\cdot\boldsymbol{s}}. \tag{4.49}$$

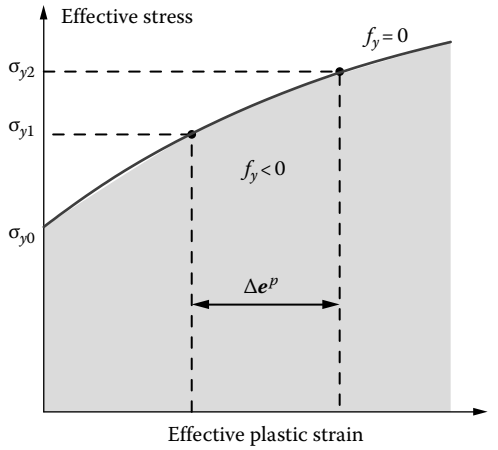


FIGURE 4.12 Generic plasticity function. The plastic strains can increase, but they appear only if $f_y = \sigma - \sigma_y = 0$. The material response has to satisfy the conditions $f_y \le 0$, $\varepsilon^p \ge 0$, $f_y \varepsilon^p = 0$.

Replacing 4.48 in 4.46 also shows that $\boldsymbol{e}''$ is oriented as $\boldsymbol{s}$:

$$\boldsymbol{s}(t) = \frac{2\mu}{1+2\mu\lambda}\boldsymbol{e}''(t). \tag{4.50}$$

The plastic multiplier λ is computed from 4.50 by using the tensor product:

$$a^2\bar{\sigma}^2(t) - d^2 = 0, \quad a = \frac{1}{2\mu} + \lambda, \quad d^2 = \frac{3}{2}\boldsymbol{e}''(t)\cdot\boldsymbol{e}''(t). \tag{4.51}$$

In general, the coefficient a in $(4.51)_1$ is a function of λ and of $\boldsymbol{s}$. Equation 4.51_1 can be written as

$$f(\bar{\sigma}^*) = a^2(\bar{\sigma}^*) - d^2 = 0 \tag{4.52}$$

and can be called effective stress function. When $f(\bar{\sigma}^*) = 0$, the effective stress $\bar{\sigma}^*$ coincides with $\bar{\sigma}(t)$. The evaluation of the final stress is reduced to the calculation of the solution of a scalar function, easily obtained with standard numerical approaches. This technique is very appealing especially for complex constitutive laws, including thermal or viscous effects. In fact, a simple bisection procedure may be used to solve the problem. For a bilinear material (see Figure 4.13a),

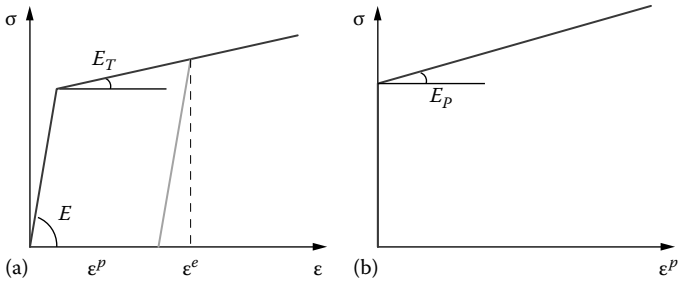


FIGURE 4.13 Bilinear elastoplastic material model. (a) Uniaxial constitutive law. (b) Rigid-plastic relationship.

Equation 4.52 becomes a linear relation, represented by the straight line of the rigid-plastic relationship $\bar{\sigma} - \bar{\varepsilon}^p$ in Figure 4.13b:

$$\Delta \bar{e}^{\,p} = \frac{\bar{\sigma}(t) - \sigma_y}{E_p}, \tag{4.53}$$

where E_P denotes the slope of the elastoplastic law:

$$E_P = \frac{EE_T}{E - E_T}$$

and E_T is the tangent modulus. From 4.53 and 4.49, one obtains

$$\bar{\sigma}(t) = \frac{2E_P d + 3\sigma_y}{E_P/\mu + 3}.$$

For simple plasticity, $E_P = 0$ and $\bar{\sigma}(t) = \sigma_y$, thus $(4.51)_1$ provides

$$\lambda = \frac{d}{\sigma_y} - \frac{1}{2\mu}.$$

The elastoplastic calculation is activated only when the material is in the condition to develop plastic strains in the time interval Δt. Therefore, the following preliminary condition has to be satisfied:

$$\bar{\sigma}^e = 2\mu\, d > \sigma_y(t_0).$$

In Figure 4.14, the solution algorithm known as elastic predictor-radial return is visualized in the deviatoric plane. The point A is the deviatoric stress at the beginning of the step, and it is assumed

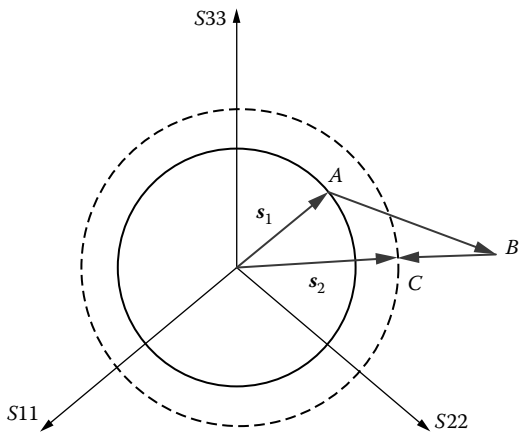


FIGURE 4.14 Geometric interpretation of the stress determination with the elastic predictor-radial return algorithm.

to lie on the plasticity surface. The position of A can be determined by sub-increments of the load step. The point B is the elastic predictor stress:

$$\boldsymbol{s}^{e}(t) = 2\mu \boldsymbol{e}''(t).$$

The corrector stress is represented by the vector CB and it is given by

$$\boldsymbol{s}^{C}(t) = 2\mu(-\Delta \boldsymbol{e}^{p}).$$

The final stress is the vector sum of the prediction stress and the corrector stress:

$$\boldsymbol{s}(t) = \boldsymbol{s}_{E}(t) + \boldsymbol{s}^{C}(t).$$

The final stress lies on the plasticity surface, hardened for the action of the plastic strains. Figure 4.14 shows that the prediction stress and the final stress are aligned in the direction of the normal to the plasticity surface $\boldsymbol{n}(t)$. This justifies the name of radial return. The plastic strain $\Delta \boldsymbol{e}^{p}$ can be written as

$$\Delta \boldsymbol{e}^{p} = \sqrt{\frac{3}{2}}\, \Delta e^{p} \boldsymbol{n}(t). \tag{4.54}$$

The interpretation of the solution algorithm gives some indication on the error of the method. The plasticity surface and the hardening rule are satisfied exactly, obviously at discrete times. Approximation is introduced only during the integration of the flow rule, where the plastic strains are estimated with a backward difference algorithm. The method is correct if the direction of the deviatoric stress $\boldsymbol{s}$ does not change during the loading step. If the normal direction $\boldsymbol{n}$ changes, though, the procedure introduces a numerical error.

One of the more restrictive requirements for the efficiency of the method is the accuracy in the calculation of the tangent stiffness matrix. To guarantee the maximum convergence rate of the Newton–Raphson algorithm, the tangent stiffness has to be consistent with the numerical scheme used to compute the stresses. This means that in terms of total stresses and strains

$$\boldsymbol{C}^{EP} = \left.\frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\varepsilon}}\right|_{t}, \tag{4.55}$$

that is, the stiffness matrix is the derivative of the stresses with respect to the strains. To include all the variables involved in the problem, the calculation of 4.55 must be done with the chain rule. For the bilinear elastoplastic law, the response depends only on the effective stress, and the chain rule imposes

$$\boldsymbol{C}^{EP} = \frac{\partial \boldsymbol{\sigma}(t)}{\partial \bar{\sigma}(t)} \frac{\partial \bar{\sigma}(t)}{\partial \boldsymbol{\varepsilon}(t)}.$$

All the differentiations have to be done in an accurate way. The final expression of the tangent stiffness matrix may result very complicated.

The previous algorithm refers to isotropic hardening, but it can be easily extended to the case of kinematic hardening, or mixed hardening, or to alternative constitutive laws (e.g., Drucker–Prager).

In such cases, more than one internal variable can be present and the consistent tangent calculation becomes more difficult.

4.6.3.2 Thermo-Elastoplasticity and Creep

The algorithm based on the effective stress function may be easily modified in order to include thermal and viscous effects, such as creep and relaxation. Temperature changes induce inelastic volumetric strains and may alter the mechanical properties of the materials.

A material model where the thermal effects modify the elastic properties, but do not modify the plastic behavior of the material, is described in the following by way of example. The material obeys isotropic hardening and von Mises plasticity condition.

The presence of creep may induce additional inelastic time-dependent strains. Following [1], Equations 4.44 and 4.45 are generalized as

$$\boldsymbol{s}(t) = 2\mu(t)[\boldsymbol{e}(t) - \boldsymbol{e}^{p}(t) - \boldsymbol{e}^{C}(t)],$$

$$\sigma_m(t) = 3K(t)[\varepsilon_m(t) - \varepsilon^{TH}(t)], \quad \varepsilon^{TH} = \alpha_m(t)[\theta(t) - \theta_{ref}],$$

where

ε^{TH} is the thermal volumetric strain

$\boldsymbol{e}^{C}$ is the deviatoric strain due to creep

The elastic constants G and K are functions of the temperature θ. Assuming that temperature is assigned during the loading history, the dependency on the temperature can be expressed as dependency on time. Thus, Equation 4.46 is written as

$$\boldsymbol{s}(t) = 2\mu(t)[\boldsymbol{e}''(t) - \Delta\boldsymbol{e}^{p} - \Delta\boldsymbol{e}^{C}].$$

All the known stresses are contained in $\boldsymbol{e}''(t)$. The incremental plastic strains $\Delta\boldsymbol{e}^{p}$ are computed as in the previous case, with the difference that $(4.51)_1$ is a function of the temperature. Two different procedures can be followed.

4.6.3.2.1 Time Hardening Procedure

The creep strains $\Delta\boldsymbol{e}^{C}$ are computed with a relation similar to the one used for the plastic strains, but using a more general procedure:

$$\Delta\boldsymbol{e}^{C} = \gamma(\tau)\boldsymbol{s}(\tau)\Delta t, \quad \boldsymbol{s}(\tau) = (1-\alpha)\,\boldsymbol{s}(t_0) + \alpha\boldsymbol{s}(t),$$

where the parameter α is chosen in the interval 0–1. The time τ falls in the interval Δt, that is, $\tau = t_0 + \alpha\Delta t$. The function $\gamma(\tau)$ is given by

$$\gamma(\tau) = \frac{3}{2}\frac{\dot{\bar{e}}^{C}(\tau)}{\bar{\sigma}(\tau)}, \tag{4.56}$$

while the effective "weighted" stress is

$$\bar{\sigma}(\tau) = (1-\alpha)\,\bar{\sigma}(t_0) + \alpha\bar{\sigma}(t). \tag{4.57}$$

The scalar function (4.56) is based on the definition of a creep law that describes the creep strain evolutions as functions of time, stresses, and temperature. A general relation, which includes several creep laws is

$$\bar{e}^{C}(t) = f_1[\bar{\sigma}(t)]\, f_2(t)\, f_3[\theta(t)] \tag{4.58}$$

and its incremental expression—which satisfies the experimental evidence and corresponds to the differentiation of the sole function f_2 with respect to time—can be reduced to

$$\Delta\bar{e}^{C}(\tau) = \Delta t f_1[\bar{\sigma}(\tau)]\, \dot{f}_2(t)\, f_3[\theta(\tau)], \quad \theta(\tau) = (1-\alpha)\theta(t_0) + \alpha\theta(t). \tag{4.59}$$

Thus, Equation 4.56 is evaluated through the expressions (4.57) and (4.59). This approach is known as "time hardening procedure."

4.6.3.2.2 Strain Hardening Procedure

When the stress is variable, it is physically more significant to express the creep strains as a function of the effective creep strain $\bar{e}^{C}$ than as a function of the time τ. From 4.58 to 4.59 with $\Delta\bar{e}^{C} = 0$, the pseudo-time τ_p is computed. The pseudo-time is defined as so as the effective creep strains equal the current strains but at a different level of the effective stress:

$$\bar{e}^{C}(t) + f_1[\bar{\sigma}(\tau)]\, f_3[\bar{\theta}(\tau)][\alpha\Delta t\dot{f}_2(\tau_p) - f_2(\tau_p)] = 0. \tag{4.60}$$

In general, Equation 4.60 has to be solved numerically. Once τ_p is known, from 4.59 the strain increment is computed. The method is called "strain hardening procedure." The difference between the time hardening method and the strain hardening method is visualized in Figure 4.15. Under cycling load, it is incumbent to account for load reversals, which renders the model more complicated.

4.6.4 Large Strain Elastoplasticity

Several formulations have been introduced in order to extend the inelastic response to finite deformations. Among them, the proposals by Green and Naghdi [8,9], Lubliner [10], Lee [11,12], Simó and Ortiz [13,14], and Ortiz and coworkers [15–17] must be quoted. In the following, an example of large strain elastoplasticity, based on the Cauchy stress $\boldsymbol{\sigma}$ and logarithmic strains $\boldsymbol{\epsilon}$, is described.

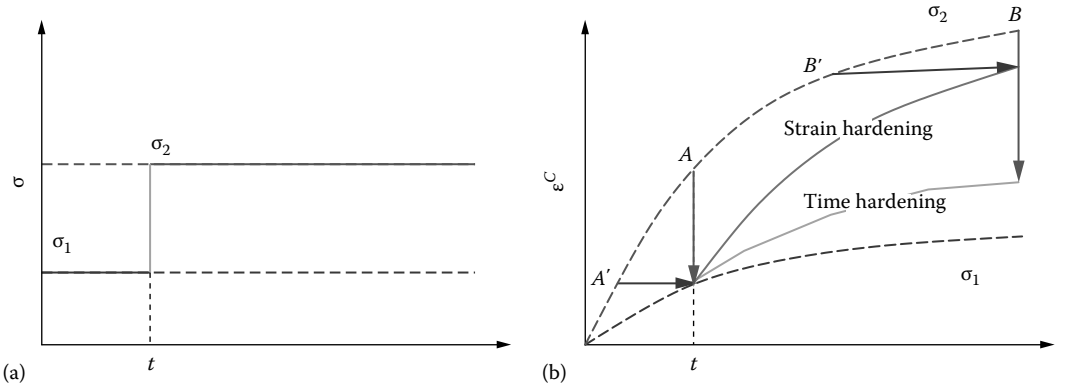


FIGURE 4.15 Creep strains in time hardening and strain hardening. (a) Stress history and (b) material response. After the stress increment in time hardening, the evolution of the creep strains follows the trend of the strains in the constant stress curve σ_2, and it is constant since the time t, curve AB. In strain hardening, the evolution of the creep strains follows the trend of the curves at constant stress σ_2 starting from the strain $\bar{e}^{C}$ at time t, curve $A'B'$.

The description of the formulation follows a few introductive remarks.

1. For any formulations, a few basic choices have to be done:
   a. The stress and strain measures (work conjugate)
   b. The characterization of the elastic behavior
   c. The flow rule
2. When the configuration changes are sufficiently small, the finite strain formulation must reduce to the infinitesimal strain formulations.
3. The previous nonlinear formulations (MN and LD-LR-SS) are expressed in terms of total stresses. This means that the equilibrium equations are written at time t and the strains are computed at time t. The numerical integration is used exclusively to compute the inelastic strain increment $\Delta\boldsymbol{\varepsilon}^p$ during the time interval Δt. The same approach can also be used in finite strain formulations, and it is especially suitable in UL formulation.
4. Alternative incremental formulations do exist. They are expressed in terms of stress rates (e.g., Jaumann tensor), strain rates, and rotation rates. All the incremental quantities (not only the inelastic strain increment) have to be integrated in the time interval Δt leading to bigger numerical errors. Thus, the accuracy of the solution is guaranteed only for small time steps. Some rate formulations bring to nonphysical mechanical responses. Recent works in the group of Bruhns and coworkers [18–20] show that incremental formulations are correct only when expressed in terms of logarithmic increments.
5. A natural approach, based on micromechanical observations of the plastic phenomena in finite strains, assumes a "hyperelastic" material description where the total deformation gradient is multiplicatively decomposed into an elastic and an inelastic part (Lee's decomposition) [11]. Such approach leads to a formulation in terms of total stresses, which can be seen as a natural extension of the small strain formulation.
6. The uniaxial constitutive law used to characterize the response is given by the uniaxial curve logarithmic strain ϵ–Cauchy stress $\boldsymbol{\sigma}$ (Figure 4.16). The plasticity condition, the flow rule, and the hardening rule have to be expressed always in terms of Cauchy stress $\boldsymbol{\sigma}$.

4.6.4.1 Governing Equations

The deformation gradient is decomposed in the product of two terms (Lee's decomposition):

$$\boldsymbol{F} = \boldsymbol{F}^e \boldsymbol{F}^p, \tag{4.61}$$

where

$\boldsymbol{F}^e$ represents the elastic part

$\boldsymbol{F}^p$ is the inelastic part

Relation (4.61) holds for any time of the deformation process.

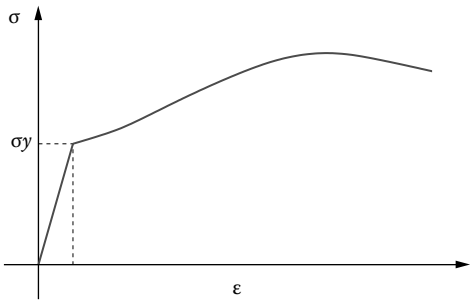


FIGURE 4.16 Uniaxial response example for elastoplasticity in finite strains.

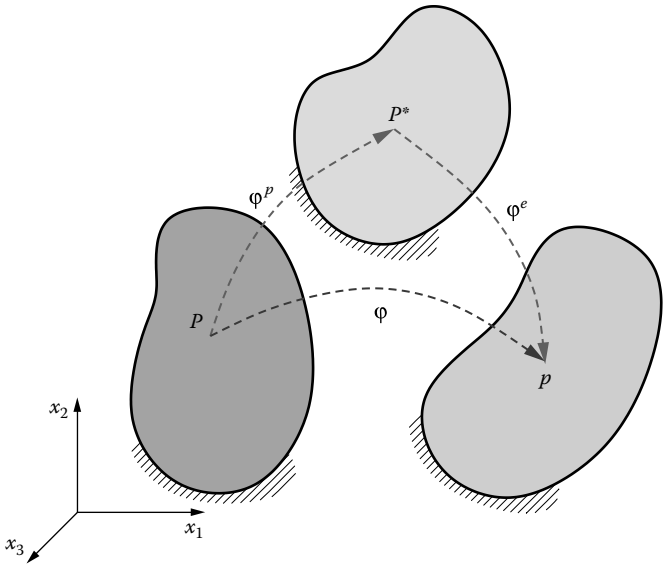


FIGURE 4.17 Multiplicative decomposition of the deformation gradient into an inelastic part $\boldsymbol{F}^p$ and into an elastic part $\boldsymbol{F}^e$.

It is convenient to think of 4.61 as two subsequent transformations (see Figure 4.17). The first transformation brings the initial configuration into a "relaxed" intermediate configuration, at the hypothetic time τ, where it is assumed that all the inelastic phenomena have already happened and the material is in a stress-free state. The fully inelastic deformation is described by the gradient $\boldsymbol{F}^p$. The second transformation, fully elastic, brings the material from the relaxed configuration to the final one. The deformation in the second step is measured by the gradient $\boldsymbol{F}^e$.

According to the experimental evidence, it makes sense to assume that the plastic deformations happen at constant volume. Therefore, the plastic Jacobian J^p can be assumed to be equal to the unit:

$$J^p = \det(\boldsymbol{F}^p) = 1, \quad J = \det(\boldsymbol{F}) = J^p J^e = \det(\boldsymbol{F}^e) = \frac{\rho}{\rho^0} > 0.$$

Equation 4.61 can be written as

$$\boldsymbol{F}^e = \boldsymbol{F}\,(\boldsymbol{F}^p)^{-1}.$$

The velocity gradient $\boldsymbol{L}$ decomposes into the sum of two terms:

$$\boldsymbol{L} = \dot{\boldsymbol{F}}(\boldsymbol{F})^{-1} = \dot{\boldsymbol{F}}^e(\boldsymbol{F}^e)^{-1} + \boldsymbol{F}^e\dot{\boldsymbol{F}}^p(\boldsymbol{F}^p)^{-1}(\boldsymbol{F}^e)^{-1} = \boldsymbol{L}^e + \boldsymbol{L}^p. \tag{4.62}$$

In order to characterize the response, it is necessary to know $\boldsymbol{F}$, $\boldsymbol{F}^p$, $\boldsymbol{\sigma}$, and the yield stress $\boldsymbol{\sigma}_y$. The Cauchy stress $\boldsymbol{\sigma}$ depends on the evolutive law of the plastic strains (or plastic deformation gradient $\boldsymbol{F}^p$) and on the hardening rule, from where σ_y is evaluated.

4.6.4.1.1 Conjugate Strain and Stress Measures

The elastic part of the deformation gradient admits the polar decomposition:

$$\boldsymbol{F}^e = \boldsymbol{R}^e \boldsymbol{U}^e, \quad \boldsymbol{C}^e = \boldsymbol{F}^{eT} \boldsymbol{F}^e = (\boldsymbol{U}^e)^2 \tag{4.63}$$

and it is spontaneous to introduce the logarithmic elastic strains (Hencky tensor) $\boldsymbol{\epsilon}^e$:

$$\boldsymbol{\epsilon}^e = \log(\boldsymbol{U}^e) = \frac{1}{2} \log \boldsymbol{C}^e.$$

The logarithmic strains refer to an intermediate configuration and $(4.63)_1$ defines the rigid rotation $\boldsymbol{R}^e$. A stress measure $\bar{\boldsymbol{\sigma}}$ corresponding to such configuration is defined as

$$\bar{\sigma} = J \boldsymbol{R}^{eT} \sigma \boldsymbol{R}^e.$$

The two stress and strain measures are conjugate, that is, their power equals to the power of $\boldsymbol{\sigma}$ and $\boldsymbol{D}^e$:

$$\bar{\sigma} \cdot \boldsymbol{\epsilon}^e = J \sigma \cdot \boldsymbol{D}^e, \quad \boldsymbol{D}^e = \frac{1}{2} (\boldsymbol{L}^e + \boldsymbol{L}^{eT}).$$

The corresponding plastic velocity gradient $\bar{\boldsymbol{L}}^p$ is obtained from 4.62:

$$\bar{\boldsymbol{L}}^p = (\boldsymbol{F}^e)^{-1} \boldsymbol{L}^p \boldsymbol{F}^e = \dot{\boldsymbol{F}}^p (\boldsymbol{F}^p)^{-1}. \tag{4.64}$$

4.6.4.1.2 Elastic Stress–Strain Relationship

The stresses $\bar{\boldsymbol{\sigma}}$ are linked to the elastic logarithmic strains $\boldsymbol{\epsilon}^e$ through a linear elastic isotropic relationship. By splitting $\boldsymbol{\epsilon}^e$ into deviatoric part $\boldsymbol{e}^e$ and isotropic part ϵ_m^e, the elastic constitutive relationship reads

$$\bar{\boldsymbol{s}} = 2\mu \boldsymbol{e}^e, \quad \bar{\sigma}_m = 3K\epsilon_m^e,$$

where $\bar{\boldsymbol{s}}$ is the deviatoric part of $\bar{\boldsymbol{\sigma}}$. The hyper-elastic law is written in terms of total stresses, and this allows for the accurate description of the stress also for moderate strains.

4.6.4.1.3 Plasticity Condition

The von Mises plasticity condition (4.47) is written in terms of deviatoric stresses $\boldsymbol{s}$. The effective stress $\bar{\sigma}$ is written as a function of the stresses in the intermediate configuration $\bar{\boldsymbol{s}}$:

$$\bar{\sigma} = \sqrt{\frac{3}{2} \boldsymbol{s} \cdot \boldsymbol{s}} = J^{-1} \sqrt{\frac{3}{2} \bar{\boldsymbol{s}} \cdot \bar{\boldsymbol{s}}}, \quad \bar{\boldsymbol{s}} = J \boldsymbol{R}^{eT} \boldsymbol{s} \boldsymbol{R}^e.$$

4.6.4.1.4 Plastic Strain Evolution Rule

The normal $\bar{\boldsymbol{n}}$ to the yield function in the intermediate configuration is given by

$$\bar{\boldsymbol{n}} = \sqrt{\frac{3}{2}} \frac{\bar{\boldsymbol{s}}}{J \bar{\sigma}}.$$

The evolutive equation for the plastic strains is obtained by generalizing the corresponding equation in small strains. The plastic strain rate is defined from the plastic velocity gradient, assuming a zero plastic spin:

$$\bar{\boldsymbol{D}}^p = \frac{1}{2}\left(\bar{\boldsymbol{L}}^p + \overline{\boldsymbol{L}^p}^T\right) = \bar{\boldsymbol{L}}^p, \quad \overline{\boldsymbol{W}}^p = \frac{1}{2}\left(\bar{\boldsymbol{L}}^p - \overline{\boldsymbol{L}^p}^T\right) = \boldsymbol{0}.$$

The flow rule (4.54) is written as

$$\bar{\boldsymbol{D}}^p = \bar{\boldsymbol{L}}^p = \sqrt{\frac{3}{2}}\,\dot{\bar{e}}^p\,\bar{\boldsymbol{n}} = \lambda\,\bar{\boldsymbol{s}}, \quad \dot{\bar{e}}^p = \sqrt{\frac{2}{3}\bar{\boldsymbol{D}}^p \cdot \bar{\boldsymbol{D}}^p}.$$

This corresponds to the maximum plastic dissipation principle [10].

4.6.4.2 Solution Procedure

The differences between the quantities used in MNL and UL are

- In the hyper-elastic law, $\bar{\boldsymbol{\sigma}}$ instead of $\boldsymbol{\sigma}$
- In the hyper-elastic law, $\boldsymbol{e}^e$ instead of $\boldsymbol{\varepsilon}^e$
- In the flow rule, $\bar{\boldsymbol{D}}^p$ instead of $\boldsymbol{\varepsilon}^p$

Assuming that at each time step the final value of the deformation gradient $\boldsymbol{F}(t)$ is assigned, the solution algorithm consists in the following steps:

1. Assignment of the trial elastic step
   Trial elastic deformation gradient:

$$\boldsymbol{F}^e(\tau) = \boldsymbol{F}(t)(\boldsymbol{F}^p(t_0))^{-1}.$$

   Elastic polar decomposition:

$$\boldsymbol{F}^e(\tau) = \boldsymbol{R}^e(\tau)\boldsymbol{U}^e(\tau).$$

   Trial elastic logarithmic strains:

$$\boldsymbol{\epsilon}^e(\tau) = \log(\boldsymbol{U}^e(\tau)).$$

   Trial Cauchy stresses (deviatoric and mean):

$$\bar{\boldsymbol{s}}(\tau) = 2\mu\boldsymbol{e}^e(\tau), \quad \bar{\sigma}_m(\tau) = 3K\epsilon^e_m(\tau).$$

   Trial equivalent effective stress:

$$\bar{\sigma}(\tau) = J(t)^{-1}\sqrt{\frac{3}{2}\bar{\boldsymbol{s}}(\tau)\cdot\bar{\boldsymbol{s}}(\tau)}.$$

2. Check of the trial stress state
   If $\bar{\sigma}(\tau) < \sigma_y(t_0)$, the trial stress is elastic.
   It results:

$$\bar{\sigma}(t) = \bar{\sigma}(\tau), \quad \sigma_y(t) = \sigma_y(t_0), \quad \boldsymbol{\epsilon}^e(t) = \boldsymbol{\epsilon}^e(\tau), \quad \bar{\sigma}(t) = \bar{\sigma}(\tau).$$

   Cauchy stresses:

$$\boldsymbol{\sigma}(t) = J(t)^{-1}\, \boldsymbol{R}^e(\tau)\, \bar{\boldsymbol{\sigma}}(\tau)\, \boldsymbol{R}^e(\tau)^T.$$

   Go back to point 1.
3. Plastic step
   The effective stress function (4.52) is solved numerically obtaining the effective stress in the intermediate configuration and the plastic strains:

$$f[\bar{\sigma}(\tau)] = a^2[\bar{\sigma}(\tau)] - d^2 = 0 \Rightarrow \bar{\sigma}(t),\ \bar{e}^p(t).$$

   Thus, the following variables are computed in sequence.
   Plastic multiplier from $(4.49)_1$:

$$\lambda(t) = \frac{3}{2}\frac{\bar{e}^p(t) - \bar{e}^p(t_0)}{J(t)\,\bar{\sigma}(t)}.$$

   Deviatoric stresses in the intermediate configuration:

$$\bar{\boldsymbol{s}}(t) = \frac{\bar{\boldsymbol{s}}(\tau)}{1 + 2\mu\lambda(t)}.$$

   Total stresses in the intermediate configuration:

$$\bar{\boldsymbol{\sigma}}(t) = \bar{\boldsymbol{s}}(t) + \frac{1}{3}\mathrm{tr}\left[\boldsymbol{\sigma}(\tau)\right]\boldsymbol{I}.$$

   Cauchy stresses:

$$\boldsymbol{\sigma}(t) = J(t)^{-1}\, \boldsymbol{R}^e(\tau)\, \bar{\boldsymbol{\sigma}}(t)\, \boldsymbol{R}^e(\tau)^T.$$

   Plastic deformation gradient, by integration of 4.64 and by using 4.65:

$$\dot{\boldsymbol{F}}^p(\boldsymbol{F}^p)^{-1} = \lambda\,\bar{\boldsymbol{s}} \Rightarrow \boldsymbol{F}^p(t) = \exp\left[\lambda(t)\bar{\boldsymbol{s}}(t)\right]\boldsymbol{F}^p(t_0).$$

The description of the solution strategy is closed by a few final remarks:

1. The previous equations are also used in bidimensional analyses (plane stress or plane strain), provided that the correct reduced constitutive relationships are used.
2. It is possible to extend the procedure to structural elements, such as beams, plates, and shells, taking care of the effects related to the change in the transversal dimensions (thickness).
3. It is possible to extend the procedure to materials showing viscous and temperature-dependent behavior, by following the scheme described for the MN formulation.
4. In the case of incompressible materials, it is always preferable to use mixed formulations (i.e., also to discretize the pressure field) both in two and three dimensions, to rule out spurious locking effects.

4.7 FINITE ELEMENT SOLUTION

The equilibrium equation in weak form, written at time t, is valid for a continuum displacement field. Let us consider a body discretized into finite elements, where, besides the displacement components, the rotation are also assumed to be independent variables.

The continuum problem is approximated by linearizing the virtual work equation with respect to all the independent variables. Because of the approximation, in both TL and UL formulations, the internal work computed at the first iteration does not coincide with the external work. The difference defines the errors:

$$\mathrm{Err}_{TL}^{(1)} = W^0(t) - \int_{V^0} S_{hk}^{(1)}\, \delta E_{hk}^{(1)}\, dV^0, \tag{4.65}$$

$$\mathrm{Err}_{UL}^{(0)} = W^t(t) - \int_{V^{t(0)}} \sigma_{hk}^{(1)}\, \delta e_{hk}^{(1)}\, dV^t. \tag{4.66}$$

The right-hand sides of 4.65 and 4.66 correspond to the right-hand sides of the relations 4.31 and 4.34, except that here they are evaluated in the current configuration, by using suitable stress and strain measures. The left-hand sides of 4.65 and 4.66 represent the out-of-balance work resulting from the linearization, at the current solution.

Obviously, a good solution has a small error. To reduce the magnitude of the error it is necessary to adopt an iterative procedure, where the equations are solved several times, by using the displacements computed at the previous iteration.

In TL formulation, the iterative expression of the displacement equation becomes

$$\int_{V^0} C_{hkij}^{t(k-1)} \Delta e_{ij}^{(k)} \delta \Delta e_{hk}\, dV^0 + \int_{V^0} S_{hk}^{(k-1)}(t) \delta \Delta \eta_{hk}^{(k)}\, dV^0 = W^0(t) - \int_{V^0} S_{hk}^{(k-1)}(t) \delta \Delta e_{hk}^{(k-1)}\, dV^0, \tag{4.67}$$

where the material tangent stiffness C_{hkij}^{t} at time t is referred to the undeformed configuration at time t_0.

In UL formulations, the iterative expression becomes

$$\int_{V^{t(k-1)}} C_{hkij}^{t(k-1)} \Delta e_{ij}^{(k)} \delta \Delta e_{hk}\, dV^t + \int_{V^{t(k-1)}} \sigma_{hk}^{(k-1)}(t) \delta \Delta \eta_{hk}^{(k)}\, dV^t = W^{t(k-1)}(t) - \int_{V^{t(k-1)}} \sigma_{hk}^{(k-1)}(t) \delta \Delta e_{hk}^{(k-1)}\, dV^t, \tag{4.68}$$

where the material tangent stiffness C_{hkij}^{t} at time t is referred to the current configuration at time t.

Between k and $(k + 1)$ iteration, the displacements are updated as

$$u_h^{(k)}(t) = u_h^{(k-1)}(t) + \Delta u_h^{(k)}(t), \quad u_h^{(0)}(t) = u_h(t_0). \tag{4.69}$$

Equations 4.67 or 4.68 and 4.69 define the Newton–Raphson procedure. All the variables are updated at each iteration, and only at the first iteration the variables evaluated at the end of the previous step are accounted for.

For loads that do not change direction with deformations, the expression of the external work (4.10) can be computed simply by evaluating the value of loads at time t:

$$W^0(t) = \int_{V^0} \rho b_i(t) \delta u_i dV^0 + \int_{S_f^0} \rho^0 t_i(t) \delta u_i dS^0. \tag{4.70}$$

If isoparametric finite elements are used, the inertia can be included in the external work. Because of the identity

$$\int_{V^0} \rho^0 \ddot{u}_i(t) \delta u_i dV^0 = \int_{V^t} \rho^t \ddot{u}_i(t) \delta u_i dV^t,$$

the mass matrix can be computed only once, at the beginning, for the initial configuration of the body.

When the external work is a function of the deformations, 4.70 cannot be used. Thus, the external work can be approximated with good accuracy by using the load intensity at time t, and considering the last calculated volumes and areas.

It is important to observe that

1. For an efficient integration scheme, the effect of the unknown incremental displacements must be included into the stiffness matrix. According to the applied load, the stiffness matrix may result nonsymmetric, causing an increase of the computational cost.
2. The virtual work equations are discussed here with reference to static or dynamic calculations performed with an implicit integration scheme. If a dynamic analysis is performed with an explicit integration scheme, the variables necessary in the calculation (velocities, accelerations) are computed at the beginning of the step. The explicit method is reduced to the evaluation of the internal work, and the continuum mechanics equations become
   - Total Lagrangian:

$$\int_{V^0} S_{hk}(t_0) \delta E_{hk} dV^0 = W^0(t).$$

   - Updated Lagrangian:

$$\int_{V^t} \sigma_{hk}(t) \delta e_{hk} dV^t = W^t(t).$$

- Materially nonlinear:

$$\int_V \sigma_{hk}(t_0)\delta e_{hk} dV = W(t).$$

The inertia is included in the external loads.

4.7.1 Finite Element Linearization of the Equilibrium Equations

The derivation of the linearized virtual work equation is obtained here in detail only for the TL formulation (4.27). For the sake of simplicity, the external work W^0 is assumed to be independent of the deformations:

$$\int_{V^0} S_{ij}(t)\delta E_{ij}(t)\, dV^0 = W^0(t).$$

Denoting with a_k the generic nodal degree of freedom (displacement or rotation) of the chosen finite element discretization, and recalling that the current time step is here defined as $\Delta t = t - t_0$, the internal work is expressed through the Taylor expansion:

$$S_{ij}(t)\delta E_{ij}(t) \approx S_{ij}(t_0)\delta E_{ij}(t_0) + \left.\frac{\partial(S_{ij}\delta E_{ij})}{\partial a_k}\right|_{t_0} da_k,$$

where da_k denotes the variation of a_k. Since the deformations change with the variation a_l, it results

$$\delta E_{ij}(t_0) = \left.\frac{\partial E_{ij}}{\partial a_l}\right|_{t_0} \delta a_l.$$

The variation is taken about the configuration at time t_0 with respect to a_l. By developing the right-hand term in 4.71, one obtains

$$\left.\frac{\partial(S_{ij}\delta E_{ij})}{\partial a_k}\right|_{t_0} da_k = \left.\frac{\partial S_{ij}}{\partial a_k}\right|_{t_0} \delta E_{ij}(t_0)\, da_k + S_{ij}(t_0)\left.\frac{\partial \delta E_{ij}}{\partial a_k}\right|_{t_0} da_k = \tag{4.71}$$

$$\left.\frac{\partial S_{ij}}{\partial E_{rs}}\right|_{t_0} \left.\frac{\partial E_{rs}}{\partial a_k}\right|_{t_0} \left.\frac{\partial E_{ij}}{\partial a_l}\right|_{t_0} \delta a_l\, da_k + S_{ij}(t_0)\frac{\partial}{\partial a_k}\left(\left.\frac{\partial E_{ij}}{\partial a_l}\right|_{t_0}\right)\delta a_l da_k$$

$$= \left(C_{ijrs}\left.\frac{\partial E_{rs}}{\partial a_k}\right|_{t_0}\left.\frac{\partial E_{ij}}{\partial a_l}\right|_{t_0} + S_{ij}(t_0)\left.\frac{\partial^2 E_{ij}}{\partial a_k\, \partial a_l}\right|_{t_0}\right)\delta a_l\, da_k. \tag{4.72}$$

C_{ijrs} is the tangent stiffness matrix of the material for the configuration at time t_0. Finally, the first and second derivatives of the Green–Lagrange tensors are written as

$$\begin{aligned}\frac{\partial E_{ij}}{\partial a_k} &= \frac{1}{2}\left(\frac{\partial u_{i,j}}{\partial a_k}+\frac{\partial u_{j,i}}{\partial a_k}+\frac{\partial u_{n,i}}{\partial a_k}u_{n,j}+u_{n,i}\frac{\partial u_{n,j}}{\partial a_k}\right)\\ \frac{\partial^2 E_{ij}}{\partial a_k\,\partial a_l} &= \frac{1}{2}\left(u_{n,i}\frac{\partial^2 u_{n,j}}{\partial a_k\,\partial a_l}+u_{n,j}\frac{\partial^2 u_{n,i}}{\partial a_k\,\partial a_l}+\frac{\partial u_{n,i}}{\partial a_k}\frac{\partial u_{n,j}}{\partial a_l}+\frac{\partial u_{n,i}}{\partial a_l}\frac{\partial u_{n,j}}{\partial a_k}\right).\end{aligned} \tag{4.73}$$

If the displacement interpolation is linear in the nodal degrees of freedom, as usually happens in solid elements, the second derivatives become

$$\frac{\partial^2 E_{ij}}{\partial a_k\,\partial a_l}=\frac{\partial u_{n,i}}{\partial a_k}\frac{\partial u_{n,j}}{\partial a_l}.$$

By replacing 4.71 and 4.72 in 4.74, the TL finite element virtual work equation is obtained:

$$\left(\int_{V^0} C_{ijrs}\frac{\partial E_{rs}}{\partial a_k}\frac{\partial E_{ij}}{\partial a_l}dV^0+\int_{V^0} S_{ij}\frac{\partial^2 E_{ij}}{\partial a_k\,\partial a_l}dV^0\right)da_k\,\delta a_l=W^0(t)-\left(\int_{V^0} S_{ij}\frac{\partial E_{ij}}{\partial a_l}dV^0\right)\delta a_l. \tag{4.74}$$

When the discretization involves isoparametric solid elements, where the degrees of freedom are only the nodal displacements, Equation 4.74 can be obtained in an alternative way, starting from the linearized expression of the internal work (4.31). But when structural finite elements are used (such as beams, plates, and shells), and rotational degrees of freedom are involved, it is always better to use Equation 4.74. In fact, 4.74 includes directly the second derivatives of the displacement gradient with respect to all the degrees of freedom. In TL-linearized formulation, Equation 4.31, the second derivatives of the displacement gradient will not be appeared explicitly, but will be included in the right-hand side terms of the equations. Their effect, though, has to be considered, since they contribute to the stiffness matrix when the displacement is not linear function of the nodal variables. Finally, relation 4.74 is useful especially for the formulation of mixed problems, where static variables (e.g., the pressure) appear in the unknowns.

A similar relation is obtained for the UL approach.

4.7.2 Matrix Equation for Solid Elements

The assembling of the matrices that govern static or dynamical problems in nonlinear formulations follows closely, in the general case, the assembling of the matrices for linear finite element analysis. Once the interpolation functions of coordinates and displacements have been chosen, the virtual work theorem is invoked for each degree of freedom, and the finite element equations are obtained.

Unlike the linear case, it is mandatory to use isoparametric elements, where the interpolation functions of the displacements are the same interpolation functions used for the coordinates. In fact, the coordinates in the deformed configuration are the sum of the coordinates in the undeformed configuration and of the nodal displacements. Therefore, the use of the same interpolation functions guarantees a consistent solution procedure. The most relevant consequence is that remarks on the convergence rate, valid for the linear finite element method, can also be extended to the incremental analysis. In particular, the compatibility of the displacements on

the element boundaries, in general imposed on the initial configuration, is also guaranteed in the deformed configurations.

By restricting the focus on the linearized equations of the virtual work theorem, the interpolation of coordinates and displacements originates the matrix equations reported in the following sections. The extension of the expressions to iterative procedures is straightforward.

4.7.2.1 Materially Nonlinear

There is no difference between the deformed and the undeformed configuration. The force array $\boldsymbol{F}^{\text{int}}(t_0)$ collects the integral of the Cauchy stress $\boldsymbol{\sigma}$ at time t_0:

$$\boldsymbol{F}^{\text{int}}(t_0) = \int_V \boldsymbol{B}_L^T \sigma dV$$

and $\boldsymbol{B}_L$ is the matrix that connects the linear part of the strain array to the displacements. $\boldsymbol{K}(t_0)$ is the small strain tangent stiffness matrix t_0, which does not account for geometrical effects:

$$\boldsymbol{K}(t_0) = \int_V \boldsymbol{B}_L^T \boldsymbol{C} \boldsymbol{B}_L dV$$

and $\boldsymbol{C}$ is the constitutive tangent stiffness.

- Static analysis:

$$\boldsymbol{K}(t_0)\Delta\boldsymbol{U} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0)$$

- Implicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t) + \boldsymbol{K}(t_0)\Delta\boldsymbol{U} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0)$$

- Explicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t_0) = \boldsymbol{R}(t_0) - \boldsymbol{F}^{\text{int}}(t_0)$$

4.7.2.2 TL Formulation

The array $\boldsymbol{F}^{\text{int}}(t_0)$ collects the forces corresponding to the second Piola–Kirchhoff stress matrix $\boldsymbol{S}$ at time t_0, computed in the undeformed configuration:

$$\boldsymbol{F}^{\text{int}}(t_0) = \int_{V^0} \boldsymbol{B}_L^T \boldsymbol{S} dV^0,$$

$\boldsymbol{B}_L$ is the matrix, referred to the initial configuration, that connects the linear part of the strains to the displacements. $\boldsymbol{K}_L(t_0)$ is the small strain tangent stiffness matrix at time t_0, referred to the undeformed configuration:

$$\boldsymbol{K}_L(t_0) = \int_{V^0} \boldsymbol{B}_L^T \boldsymbol{C} \boldsymbol{B}_L dV^0,$$

$\boldsymbol{C}$ is the constitutive tangent stiffness, referred to the initial configuration. $\boldsymbol{K}_{NL}(t_0)$ is the geometrical tangent stiffness, which accounts for the geometric nonlinearities and for the initial stresses, computed in the undeformed configuration:

$$\boldsymbol{K}_{NL}(t_0) = \int_{V^0} \boldsymbol{B}_{NL}^T \boldsymbol{S} \boldsymbol{B}_{NL} dV^0,$$

$\boldsymbol{B}_{NL}$ is the matrix linking the nonlinear part of the strains to the displacements, referred to the initial configuration.

- Static analysis:

$$[\boldsymbol{K}_L(t_0) + \boldsymbol{K}_{NL}(t_0)]\Delta \boldsymbol{U} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0).$$

- Implicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t) + \left[\boldsymbol{K}_L(t_0) + \boldsymbol{K}_{NL}(t_0)\right]\Delta \boldsymbol{U} = \boldsymbol{R}(t) - \boldsymbol{F}^{\text{int}}(t_0).$$

- Explicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t_0) = \boldsymbol{R}(t_0) - \boldsymbol{F}^{\text{int}}(t_0).$$

4.7.2.3 UL Formulation

The array $\boldsymbol{F}^{\text{int}}(t_0)$ collects the contribution of the Cauchy stress $\boldsymbol{\sigma}$ at time t_0, computed with reference to the deformed configuration at time t_0:

$$\boldsymbol{F}^{\text{int}}(t_0) = \int_{V^t} \boldsymbol{B}_L^T \boldsymbol{\sigma} dV^t.$$

$\boldsymbol{B}_L$ is the matrix that relates the linear part of the strains to the displacements, referred to the time t_0. $\boldsymbol{K}_L(t_0)$ is the small strain tangent stiffness at time t_0, computed with reference to the deformed configuration:

$$\boldsymbol{K}_L(t_0) = \int_{V^t} \boldsymbol{B}_L^T \boldsymbol{C} \boldsymbol{B}_L dV.$$

$\boldsymbol{C}$ is the constitutive stiffness matrix, referred to the deformed configuration at time t_0. $\boldsymbol{K}_{NL}(t_0)$ is the tangent stiffness matrix, which accounts for the geometric nonlinearities and initial stresses, computed with reference at the deformed configuration at time t_0:

$$\boldsymbol{K}_{NL}(t_0) = \int_{V^t} \boldsymbol{B}_{NL}^T \boldsymbol{\sigma} \boldsymbol{B}_{NL} dV^t.$$

$\boldsymbol{B}_{NL}$ is the matrix, which links the nonlinear part of the strains to the displacements.

- Static analysis:

$$[\boldsymbol{K}_L(t_0)+\boldsymbol{K}_{NL}(t_0)]\,\Delta\boldsymbol{U}=\boldsymbol{R}(t)-\boldsymbol{F}^{\text{int}}(t_0).$$

- Implicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t)+[\boldsymbol{K}_L(t_0)+\boldsymbol{K}_{NL}(t_0)]\,\Delta\boldsymbol{U}=\boldsymbol{R}(t)-\boldsymbol{F}^{\text{int}}(t_0).$$

- Explicit dynamic analysis:

$$\boldsymbol{M}\ddot{\boldsymbol{U}}(t_0)=\boldsymbol{R}(t_0)-\boldsymbol{F}^{\text{int}}(t_0).$$

In the previous equations, $\boldsymbol{M}$ is the mass matrix, independent of the strain; $\Delta\boldsymbol{U}$ is the incremental nodal displacement array; $\ddot{\boldsymbol{U}}(t_0)$ and $\ddot{\boldsymbol{U}}(t)$ are the nodal acceleration arrays, at times t_0 and t, respectively; $\boldsymbol{R}(t_0)$ and $\boldsymbol{R}(t)$ are the external nodal load arrays, at times t_0 and t, respectively.

4.8 SPECIAL PROBLEMS

For the solution of standard statics and dynamics problems, two restrictive assumptions have been made: (1) the compatibility of the displacement field, described by regular functions and (2) the preservation of the body boundaries, with a clear distinction between the displacement boundary and the traction boundary. When one of these assumptions is missing, two new kinds of mechanical problems have to be considered. Fracture mechanics deals with problems where displacement compatibility fails, while contact mechanics deals with boundaries where both displacements and tractions are unknowns.

4.8.1 Nonlinear Fracture Mechanics

The cohesive approach to fracture mechanics [21,22] describes a crack as the progressive separation of two originally coincident surfaces S^+ and S^- (see Figure 4.18). Every point P^+ belonging to S^+ has a corresponding point P^- on S^-. It is possible to define a mean surface, equidistant from the two original surfaces, where static and kinematics quantities can be defined.

The two surfaces move away as consequence of the displacement $\boldsymbol{u}^+$ of $\boldsymbol{u}^-$ of the two points P^+ and P^-. The displacement jump is the difference:

$$[[\boldsymbol{u}]]=\boldsymbol{u}^+-\boldsymbol{u}^-.$$

The separation of the two points is contrasted by the cohesive tractions $\boldsymbol{t}^c$. The cohesive tractions are related to the displacement jump through the cohesive law, or interface law, characterized by a progressive softening of the response:

$$\boldsymbol{t}^c=f\,[[\boldsymbol{u}]]. \tag{4.75}$$

The cohesive law has to be able to describe the different types of separation, that is, the opening mode from the shear modes. Therefore, the expression of 4.75 can become quite complicated. A way to simplify the analytical treatment of the cohesive law is to define effective scalar quantities, by weighting the components of the displacement jump in the normal and tangential direction with respect to the cohesive surface [16,23]. An example of cohesive law expressed in terms of effective variables is reported in Figure 4.19a.

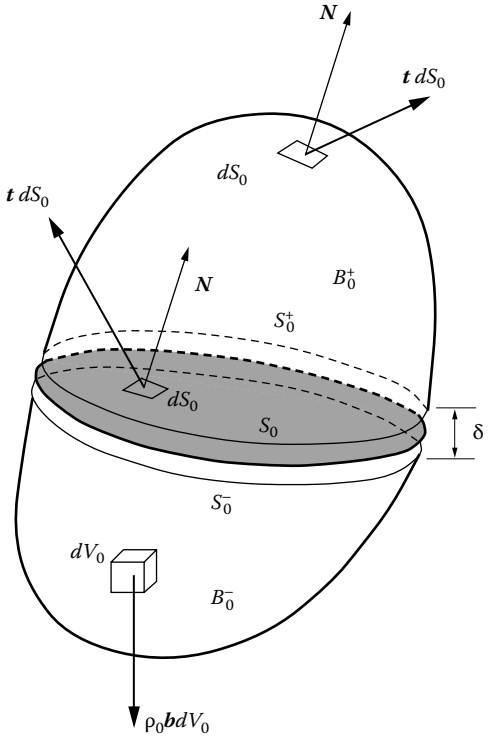


FIGURE 4.18 Cohesive surface across a continuum.

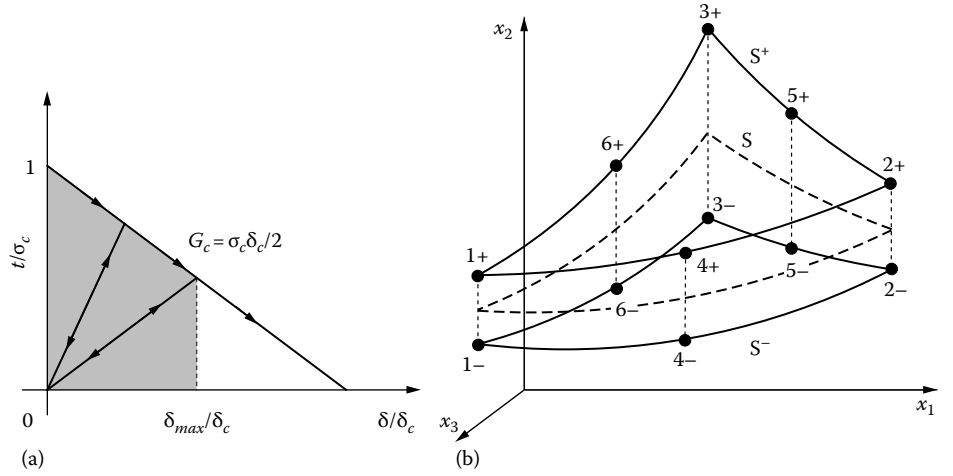


FIGURE 4.19 (a) Irreversible linear cohesive law. (b) Interface finite element, compatible with the discretization of the continuum into quadric tetrahedra.

The presence of a cohesive surface introduces a new term in the virtual work equation, that is, the work done by the cohesive tractions for the displacement jump:

$$\int_{V(t)} \sigma_{ij}(t)\,\delta e_{ij}\,dV = \int_{V(t)} \rho(t) b_i(t)\,\delta u_i\,dV + \int_{S_f(t)} t_i(t)\,\delta u_i\,dS + \int_{S_c(t)} t_i^c(t)\,\delta \llbracket u \rrbracket_i\,dS.$$

Cohesive models are especially appealing for the finite element applications. With a simple modification of a standard code, it is possible to apply the cohesive laws as mixed boundary conditions.

Alternatively, it is possible to introduce interface finite elements, made of two surfaces coinciding with the boundary surface of two adjacent solid elements (Figure 4.19).

The cohesive law governs the response of the material when fracture occurs, and defines the value of the traction $\boldsymbol{t}_c$ on the cohesive surface. The corresponding nodal forces are evaluated by integrating the cohesive tractions on the cohesive surface:

$$\boldsymbol{F}^{\text{int}}(t)^{\pm} = \mp \int_{S_0} \boldsymbol{t}_c \boldsymbol{N} dS_0, \tag{4.76}$$

where $\boldsymbol{N}$ is the shape function matrix. The tangent stiffness matrix is computed through a consistent linearization of 4.76:

$$\boldsymbol{K}(t)^{\pm\pm} = \mp \mp \frac{\partial \boldsymbol{F}^{\text{int}}(t)^{\pm}}{\partial \boldsymbol{x}^{\pm}}.$$

The tangent stiffness matrix may result nonsymmetric if geometrical nonlinearities are considered.

4.8.2 Contact Mechanics

The formulation of the contact problem, from the small strain frictionless contact to the large strain frictional contact, is always the same, although the difficulties related to the numerical solution may be very different.

Contact between two or more bodies is observed when their motion is somehow constrained. The bodies cannot assume all the possible configurations, but geometrical constraints impede their motion partially or completely. Contact involves the boundaries of the bodies (Figure 4.20), and the constraint reactions become surface normal and tangent tractions.

Figure 4.21a shows the so-called Signorini's condition that relates the normal relative displacements $u_N = [\![\boldsymbol{u}^c]\!] \cdot \boldsymbol{n}$ to the compressive tractions $T_N = \boldsymbol{t}^c \cdot \boldsymbol{n}$ when two bodies are in contact:

$$u_N \geq 0, \quad T_N \leq 0, \quad u_N T_N = 0.$$

Figure 4.21 shows the condition on the tangential velocities $\dot{u}_S$ that are observed in the direction opposite to the resisting tangential forces, only when the force reaches a limit value (Coulomb law):

$$\dot{u}_S = 0 \quad \text{if } |T_S| < \mu |T_N|, \quad \text{sign}(\dot{u}_S) = \text{sign}(T_S) \quad \text{if } |T_S| = \mu |T_N|,$$

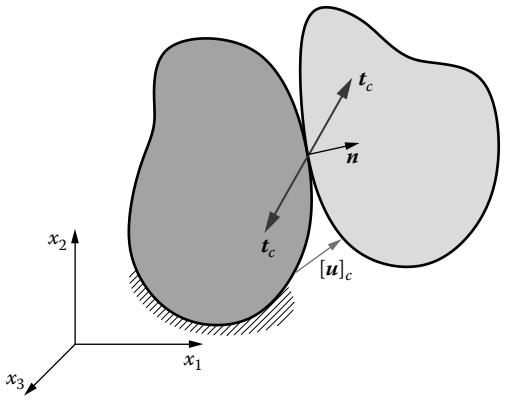


FIGURE 4.20 Smooth contact between two bodies, definition of the normal $\boldsymbol{n}$ to one of the two surfaces and of the contact tractions $\boldsymbol{t}^c$. In the contact surface, the displacement jump $[\![\boldsymbol{u}^c]\!]$ is null.

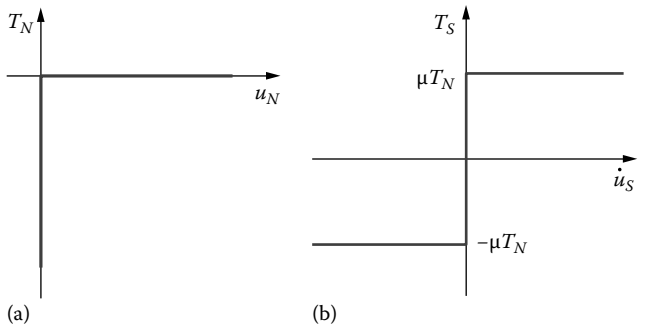


FIGURE 4.21 Frictional contact conditions: (a) Signorini's condition for the unilateral contact and (b) tangent velocity condition.

where μ is the friction coeffcient. By nature, contact is non-smooth and a rigorous treatment of the problem requires to use the non-smooth analysis [24].

A classic approach, based on the definition of a gap function between the two surfaces in potential contact, is presented here. The approach relies on the existence of a smooth contact surface [1].

For a system of N bodies in motion, the virtual work theorem is written by, accounting for the contribution of all the bodies, including the work of the contact forces. The weak form of the equilibrium equation is

$$\sum_{\text{Bodies}} \left\{ \int_{V(t)} \sigma_{ij}(t)\,\delta e_{ij}\, dV \right\} = \sum_{\text{Bodies}} \left\{ \int_{V(t)} \rho(t) b_i(t)\,\delta u_i\, dV + \int_{S_f(t)} t_i(t)\delta u_i\, dS \right\} + \sum_{\text{Contacts}} \int_{S_c(t)} t_i^c(t)\,\delta u_i^c\, dS, \tag{4.77}$$

where

S_c is the contact surface
$t_i^c(t)$ is the contact reaction
δu_i^c is the virtual displacement on the contact surfaces

For the sake of simplicity, let us consider two bodies A and B in contact (or two parts of the same body, in the case of self-contact). The contact reactions must be equal and opposite $\boldsymbol{t}^{AB} = -\boldsymbol{t}^{BA}$. Denoting with $S^{AB}(t)$ the total contact surface, the virtual contact work can be written as

$$\int_{S^A(t)} t_i^{AB}(t)\,\delta u_i^A\, dS + \int_{S^B(t)} t_i^{BA}(t)\,\delta u_i^B\, dS = \int_{S^{AB}(t)} t_i^{AB}(t)\,\delta \left[\!\left[u_i^{AB} \right]\!\right] dS.$$

The contact surface is defined always as a pair of surfaces, with possibly a different extension. One of the two surfaces (A) is called arbitrarily as contactor and the second (B) as target. Denoting the normal to the target surface with $\boldsymbol{n}$, $\boldsymbol{s}$ is a unit vector on the tangent plane that defines with $\boldsymbol{n}$ an orthonormal basis. The contact forces are split in normal and tangential components:

$$T_N = \boldsymbol{t}^{AB}(t)\cdot\boldsymbol{n}, \quad T_S = \boldsymbol{t}^{AB}(t)\cdot\boldsymbol{s}.$$

The pair $\boldsymbol{n}$, $\boldsymbol{s}$ is computed at a generic point $\boldsymbol{x}$ on the contactor surface. On the target surface, the point $\boldsymbol{y}^*(\boldsymbol{x}, t)$, closest to $\boldsymbol{x}$, is identified:

$$\|\boldsymbol{x} - \boldsymbol{y}^*(\boldsymbol{x}, t)\| = \min_{\boldsymbol{y}\in S^B}\{\|\boldsymbol{x} - \boldsymbol{y}\|\}.$$

The distance between $\boldsymbol{x}$ and $\boldsymbol{y}^*$ is given by the component of the vector $\boldsymbol{x} - \boldsymbol{y}^*$ on the normal $\boldsymbol{n}$:

$$u_N = g(\boldsymbol{x}, t) = (\boldsymbol{x} - \boldsymbol{y}^*(\boldsymbol{x}, t))\cdot \boldsymbol{n}$$

and g is usually called gap function. In presence of Coulomb friction, the tangential velocity $\dot{u}_S(\boldsymbol{x}, t)$ is introduced:

$$\dot{u}_S(\boldsymbol{x}, t) = [\dot{\boldsymbol{u}}(\boldsymbol{x}, t) - \dot{\boldsymbol{u}}(\boldsymbol{y}^*, t)]\cdot \boldsymbol{s}.$$

The solution of the frictional contact problem is obtained, therefore, from the solution of the virtual work equation written in the form (4.77), under the constraints

$$\begin{gathered} u_N \geq 0 \\ T_N \leq 0 \\ u_N T_N = 0 \\ |T_S| \leq |T_N| \\ \dot{u}_S = 0 \quad \text{if } |T_S| < \mu |T_N| \\ \operatorname{sign}(\dot{u}_S) = \operatorname{sign}(T_S) \quad \text{if } |T_S| = \mu |T_N|. \end{gathered} \tag{4.78}$$

In dynamic analyses, the inertia forces have to be included between the body forces, and additional kinematic conditions on accelerations and velocities have to be considered. In general, the typically adopted time integration schemes (trapezoidal rule) are not able to satisfy the contact conditions; therefore, they have to be enforced individually during the step-by-step solutions.

Many procedures have been proposed in the literature for the numerical solution of contact. Most of them are based on penalty approaches or on the introduction of Lagrange multipliers to account for the contact constraints (see, for example, [1,25,26]).

The formulation of frictional contact problems leads in general to nonsymmetric equation systems. This difficulty can be avoided if dissipative effects of friction are disregarded. Also, in the case of frictionless contact, though, the solution algorithms may show a considerable reduction of the convergence rate. Even with consistent approaches, the convergence ability can be spoiled by the choice of an inaccurate prediction step. A wide step, for example, may miss completely the detection of the contact.

Algorithm bases on the definition of gap functions are in general weak in the cases of corners, where the normal is not defined. In such cases, the stiffness matrix may intrinsically not consistent and introduces numerical instabilities. To avoid such issues, regularization methods for the contact surfaces have been proposed. Some of them are based on the introduction of smooth interpolation functions, definition of the distance between interior points of the bodies (pinball method), or, finally, the direct use of the non-smooth mechanics [25,26].

4.9 CONCLUSIONS

The finite element method is used to solve the mathematical problems that model physical problems in engineering analysis and design. A linear model can be adopted for many practical mechanical applications, when the only difficulty is the complex geometry of the system.

Other common problems in mechanics require to adopt nonlinear models. Nonlinearities are mainly due to the behavior of materials, the magnitude of the displacements, or the presence of unilateral constraints. The analysis of such problems can be pursued only with a nonlinear finite element approach.

Obviously, the presentation in the previous sections neglected many mechanical problems that also require to use nonlinear analysis, for example, the evaluation of the critical load, or the attainment of buckling conditions in slender structures [27]. According to the magnitude of the displacements, critical loads can be determined by computing the eigenvalues of the stiffness matrix (second-order analysis), or, in a fully nonlinear approach, through a step-by-step analysis with increasing loads. The latter approach allows for tracking the post-critical conditions.

Additionally, the presentation did not discuss the numerical difficulties in the iterative solution of the nonlinear systems. In proximity of critical conditions, the Newton–Raphson procedure may fail, and alternative solvers need to be adopted, such as the arc-length method [28], the conjugate gradient method [29], or the dynamic relaxation method [30–32].

Finally, a non-negligible consequence of finite kinematics is the geometrical distortion of finite elements. This effect may reduce dramatically the accuracy of the solution. Therefore, a continuous check on the element distortion may be necessary to obtain meaningful solutions. Adaptive remeshing techniques, able to replace distorted elements with patches of new elements, are becoming more and more popular and efficient [33].

REFERENCES

1. K. J. Bathe. *Finite Element Procedures*. Prentice-Hall, New York, 1996.
2. L. E. Malvern. *Introduction to the Mechanics of a Continuous Medium*. Prentice-Hall, Englewood Cliffs, NJ, 1969.
3. Y. C. Fung. *First Course in Continuum Mechanics*. Pearson Education, New York, 1993.
4. A. E. Green and W. Zerna. *Theoretical Elasticity*. Oxford University Press, London, 1968.
5. R. W. Ogden. *Non-linear Elastic Deformations*. Halsted Press, New York, 1984.
6. R. W. Ogden. *Non-linear Elastic Deformations*. Dover Publications, New York, 1997.
7. L. Corradi dell'Acqua. *Meccanica delle Strutture. I. Il Comportamento dei Mezzi Continui*. McGraw-Hill, Milano, 1992.
8. A. E. Green and P. M. Naghdi. A general theory of an elasto-plastic continuum. *Archive for Rational Mechanics and Analysis*, 18:251–281, 1965.
9. A. E. Green and P. M. Naghdi. Some remarks in elastic-plastic deformations at finite strains. *International Journal of Engineering Science*, 9:1219–1229, 1971.
10. J. Lubliner. *Plasticity Theory*. Prentice Hall, Upper Saddle River, NJ, 1998.
11. E. H. Lee. Elastic–plastic deformations at finite strains. *Journal of Applied Mechanics*, 26:1–6, 1969.
12. E. H. Lee. Some comments in elastic-plastic analysis. *International Journal of Solids and Structures*, 17:859–872, 1981.
13. J. C. Simó and M. Ortiz. A unified approach to finite deformation elasto-plastic analysis based on the use of hyperelastic constitutive tensors. *Computer Methods in Applied Mechanics and Engineering*, 49:221–245, 1985.
14. J. C. Simó. Algorithms for static and dynamic multiplicative plasticity that preserve the classical return mapping schemes of the infinitesimal theory. *Computer Methods in Applied Mechanics and Engineering*, 98:41–104, 1992.
15. A. Cuitino and M. Ortiz. A material-independent method for extending stress update algorithms from small-strain plasticity to finite plasticity with multiplicative cinematics. *Engineering Computations*, 9:437–451, 1992.

16. M. Ortiz and L. Stainier. The variational formulation of viscoplastic constitutive updates. *Computer Methods in Applied Mechanics and Engineering*, 171:419–444, 1999.
17. M. Ortiz, R. A. Radovitzky, and E. A. Repetto. The computation of the exponential and logarithmic mapping and their first and second linearization. *International Journal for Numerical Methods in Engineering*, 52:1431–1441, 2001.
18. H. Xiao, O. T. Bruhns, and A. Meyers. Hypo-elasticity model based upon the logarithmic stress rate. *Journal of Elasticity*, 47:51–68, 1997.
19. O. T. Bruhns, H. Xiao, and A. Meyers. Self-consistent Eulerian rate type elastoplasticity models based upon the logarithmic stress rate. *International Journal of Plasticity*, 15:479–420, 1999.
20. H. Xiao and L. S. Chen. Hencky's elasticity model and linear stress–strain relations in isotropic finite hyperelasticity. *Acta Mechanica*, 157:51–60, 2002.
21. D. S. Dugdale. Yielding of steel sheets containing slits. Journal of the Mechanics and Physics of Solids, 8:100–104, 1960.
22. G. I. Barenblatt. The mathematical theory of equilibrium of cracks in brittle fracture. *Advances in Applied Mechanics*, 7:55–129, 1962.
23. G. T. Camacho and M. Ortiz. Computational modelling of impact damage in brittle materials. *International Journal of Solids and Structures*, 33(20–22):2899–2938, 1996.
24. F. H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983.
25. C. Kane, E. A. Repetto, M. Ortiz, and J. E. Marsden. Finite element analysis of nonsmooth contact. *Computer Methods in Applied Mechanics and Engineering*, 180:1–26, 1999.
26. A. Pandolfi, C. Kane, M. Ortiz, and J. E. Marsden. Time discretized variational formulation of nonsmooth frictional contact. *International Journal for Numerical Methods in Engineering*, 53:1801–1829, 2002.
27. S. P. Timoshenko and J. M. Gere. *Theory of Elastic Stability*. McGraw-Hill, New York, 1961.
28. E. Riks. An incremental approach to the solution of snapping and buckling problems. *International Journal of Solids and Structures*, 15:529–551, 1979.
29. M. R. Estenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 19:410–436, 1952.
30. J. R. H. Otter. Computations for prestressed concrete reactor pressure vessels using dynamic relaxation. Nuclear and Structural Engineering, 1:61–75, 1965.
31. J. R. H. Otter. Dynamic relaxation. *Proceedings of the Institution of Civil Engineers*, 35:723–750, 1966.
32. D. R. Oakley and N. F. Knight, Jr. Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures. *Computer Methods in Applied Mechanics and Engineering*, 126:67–89, 1995.
33. T. Strouboulis, K. Copps, and I. Babuska. The generalized finite element method. *Computer Methods in Applied Mechanics and Engineering*, 190:4081–4193, 2001.

5 Nonlinear Vehicle Dynamics

Hans True

CONTENTS

5.1 Introduction 121
5.2 Formulation of a Problem in Vehicle Dynamics 123
5.3 On the Geometry of Nonlinear Autonomous Dynamic Systems with Constant Parameters 125
5.3.1 Some Definitions 125
5.3.2 Stability Concepts 126
5.3.3 Equilibrium Solutions 128
5.3.3.1 Equilibrium Points 129
5.3.3.2 Cycles 129
5.3.3.3 Chaos 133
5.4 Parameter-Dependent Systems and Bifurcations 135
5.5 Center Manifold Reduction Theorem 138
5.6 On Stochastic Dynamics and the Influence of Noise 138
5.7 On Non-Smooth Systems 141
5.8 On the Numerical Analysis of Vehicle Dynamic Systems 146
5.8.1 Numerical Solution of Smooth Problems 146
5.8.2 Numerical Solution across a Discontinuity 149
5.8.3 A Way to Compute Stable Branches of Vehicle Dynamic Problems Numerically 153
5.9 A Guide to Vehicle System Dynamics 155
Acknowledgments 156
References 156

5.1 INTRODUCTION

In general, all kinematic and constitutive relations in vehicle system dynamics are nonlinear. The modeling of such systems therefore leads to nonlinear dynamic systems, which must be investigated with respect to their properties and evolution.

It is characteristic for vehicle dynamic systems that they contain elements, which dissipate energy. Therefore, we consider only dissipative and nonconservative systems in this chapter.

In vehicle system dynamics, the dynamic problem is most often first to find the stationary state(s), second to search for other steady states, third to determine their stability, and finally to determine the behavior of the vehicle when changes between steady states, for example, by braking a car or negotiating a curve, occur. Therefore, the fundamental problem is to find the stationary and other steady states. *The analysis of a nonlinear dynamic problem must therefore always begin with a search for all the steady states depending on the parameters in the problem by solution of the full nonlinear problem.* In an appropriate frame at least one of the stationary states can often be found easily, but most often numerical methods must be used. The use of numerical methods will be discussed in Section 5.8. If multiple steady states exist, then they fortunately very often develop through so-called bifurcations or branching off from other steady states, when a control parameter in the problem is varied. This fact offers us a possibility to find multiple states by tracing already found steady-state solutions by varying the control parameter through the values where bifurcations take place and

then change to follow the new solution from the bifurcation in dependence on the control parameter. The fundamental steps in the solution of a nonlinear dynamic problem therefore consist of finding at least one stationary solution in a sufficiently large parameter interval and determine its stability in dependence on the parameter. Then find the bifurcation points where different steady solutions will be found. The new solutions and their stability must then be determined in dependence on the control parameter. The new solutions may create new bifurcation points, which must be found, and the game starts all over in these new bifurcation points. Unfortunately, there is no guarantee that *all* the steady-state solutions will be determined in this way, but experience shows that the method is very efficient. Good engineering know-how is still needed to judge whether all possible states have been found or it is necessary to use other strategies to find possible missing steady states.

Until the 1960s, the mathematics allowed us only to determine certain properties of the solutions of some nonlinear problems. In general, it was impossible to calculate the solutions of any but the simplest nonlinear problem. Various so-called asymptotic methods that expressed the solution in a series expansion that might converge were often very effective for finding approximate solutions, but the validity of the solutions was limited. Only the modern computer allows us to find solutions of complicated nonlinear problems that are valid for all values of the parameters and in the entire range of interest of the state variables.

An important property of the solution of a dynamic problem is the stability of its solutions. The stability investigation of a given solution leads to the determination of the eigenvalues with the largest real part of a *linear problem*. Early experience with nonlinear dynamic systems showed that the systems tend to a different kind of motion, when the original solution loses its stability—if it does not blow up. This new motion was therefore called an *instability* by the engineers and physicists. It is very unfortunate. The new motion is measurable and visible and it must therefore be stable. It is just another stable solution of the nonlinear problem, and its determination is a *problem of existence* of multiple solutions of nonlinear problems. The use of the term "instability" in this case is misleading—first of all because the motion called "instability" is stable—and also because the calculation of the solution is *not* a mathematical problem of stability but a problem of existence.

A nonlinear dynamic problem may have several different stable steady-state solutions—even though it is dissipative (see an example of multiple steady-state solutions in Figure 5.1). Each of them is uniquely determined by its own set of initial conditions. It means that we can relate a set of initial conditions uniquely to a stable steady-state solution of a nonlinear dynamic problem. Such a set of initial conditions is called a *basin of attraction*. In contrast, the solution of a *linear* dissipative dynamic system will always tend to a unique solution independently of the initial conditions. This fact has an important implication. If the dynamic problem has multiple stable steady states and the corresponding physical vehicle in the real world is subject to uncontrolled initial conditions, then finite probabilities exist for the vehicle to tend to anyone of the stable steady states. In other words, when several stable steady-state solutions of a dissipative problem coexist then the vehicle is *unstable to finite disturbances*.

Nonlinear mechanics deals with the formulation or modeling of processes, while nonlinear dynamics is a mathematical discipline. The topics therefore are different. We shall deal with nonlinear and *deterministic* dynamics in this chapter. Some basic information on *stochastic* dynamics is presented in Section 5.6. In nonlinear dynamics, the *existence* of solutions is the primary object of analysis in contrast to *stability*, which comes second.

In the following we shall introduce the basic notions of a *state space* and *trajectories*, because they lead to a visualization of the solutions of dynamic problems, their uniqueness (or lack of same), and stability. We shall also present the conditions under which a linearization of a nonlinear operator (problem) is permitted and indicate how the solution of the linearized problem can give valuable information about the local properties of the nonlinear problem. We shall define and examine stationary and other steady-state solutions, their stability, and bifurcations. Finally, properties of solutions that cannot exist in linear systems—such as chaos—will be described and their effects are briefly discussed.

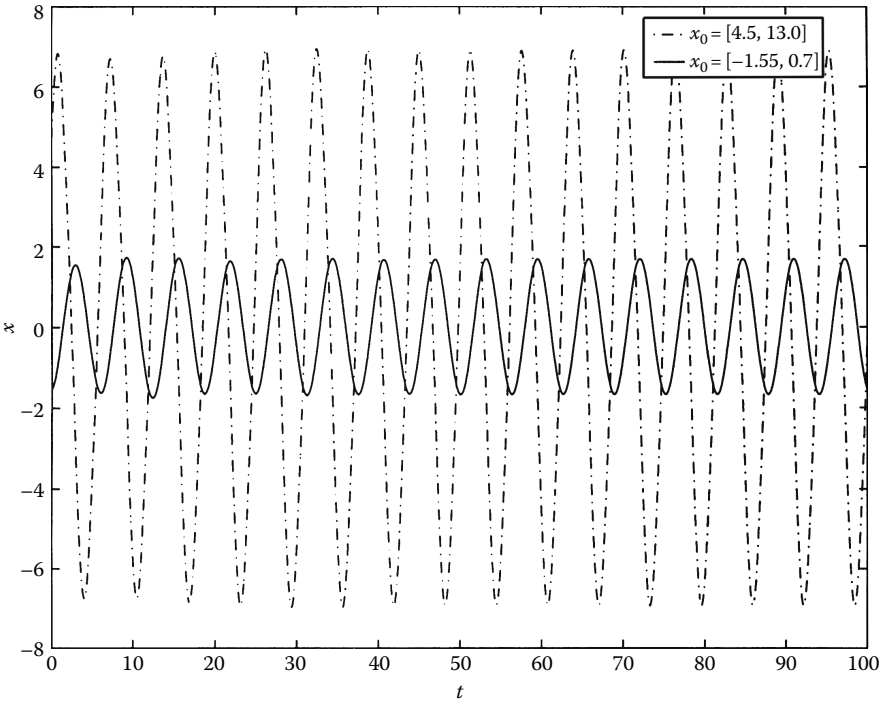


FIGURE 5.1 The two solutions of the initial value problem.

5.2 FORMULATION OF A PROBLEM IN VEHICLE DYNAMICS

The problems in theoretical vehicle dynamics are formulated as nonlinear dynamic multi-body problems. The rolling wheel–ground contact as well as the more rarely applied sliding contact are the most important features of a vehicle dynamic system, and the constitutive laws for these contacts are nonlinear. In a multi-body system the single bodies are interconnected, and several of the connections—for example, springs and dampers—obey nonlinear constitutive laws. Guides introduce constraints, which often are formulated by algebraic or transcendental equations, so the problems in vehicle dynamics must often be formulated as a system of differential-algebraic equations (DAEs). Bumper stops and dry friction contact are important nonlinearities that introduce discontinuities. The discontinuities violate the assumptions of the mathematical theorems, which form the basis of many methods that are applied for the solution of systems of DAEs.

In vehicle dynamics, the analysis of the problem is most often formulated as an initial value problem for a system of ordinary differential equations with time t as the single independent variable. The system of differential equations may be combined with algebraic or transcendental equations that express certain kinematic constraints, for example, the geometric contact between a wheelset and the rails. The constraint equations are in general nonlinear, and they must be solved together with the system of differential equations as one system of DAEs. Boundary value problems for partial differential equations or equations with time lag may, however, also enter the dynamic system, but they will not be considered in this brief survey.

The mathematical properties of nonlinear systems are fundamentally different from the mathematical properties of linear systems, and the nonlinear systems must be treated accordingly. An important difference between the properties of linear and nonlinear systems is that the principle of superposition does not apply to nonlinear systems. The principle of superposition is the basis of integral transform solution methods. Another difference is the effect of the initial conditions on the

steady-state solution of an initial value problem. A linear problem has only one steady-state solution independent of the initial conditions, but a nonlinear problem may have more steady-state solutions, where each one depends uniquely on the applied initial conditions. The existence of more steady-state solutions of a nonlinear problem is demonstrated in the following example of a nonlinear initial value problem [1].

Consider the nonlinear differential equation:

$$2.56\mathrm{d}^2x/\mathrm{d}t^2 + 0.32\mathrm{d}x/\mathrm{d}t + x + 0.05x^3 = 2.50\cos t$$

with the two initial conditions:

$$(x, \mathrm{d}x/\mathrm{d}t) = (4.5, 13.0) \quad \text{and} \quad (x, \mathrm{d}x/\mathrm{d}t) = (-1.55, 0.70) \quad \text{for } t = 0.$$

The initial value problems are easily solved with MATLAB®, and the results are shown in Figure 5.1. It is easily seen that although the excitation in both cases is a periodic function with the amplitude 2.50 and the system is damped, then there exists a steady-state large amplitude solution to the differential equation in addition to the expected steady-state solution with amplitude less than two. This is an example of *coexisting stable steady-state solutions*—called attractors, a term that will be defined later. When the corresponding linear problem is solved without the x^3 term, one and the same low amplitude steady-state periodic solution exists for both sets of initial conditions.

A dynamic system with two attractors is called *bistable*. If it has more than two attractors, it is called *multi-stable*.

Furthermore, oscillation frequencies are amplitude-dependent in nonlinear dynamics. As a consequence hereof, *the concept of resonance such as it is used in linear theory must be reconsidered and revised*.

The problems in vehicle dynamics are parameter-dependent. Examples of parameters are a wheel base, a weather-dependent friction coefficient, and the speed of the vehicle.

The state of the system is defined as the set of positions and velocities of each and every element of the vehicle in a suitable coordinate system. Both the positions and the velocities are treated as dependent variables in the dynamic system with time t as the independent variable. The dimension N of the state space equals the number of dependent variables of the dynamic system. The mathematical problem consists of finding the state of the system under external action as a function of the time and in dependence on the parameters. In Section 5.3, we shall start with the assumption that the parameters are all constant.

If time t does not enter the dynamic system explicitly, the system is called *autonomous*. If time t enters the dynamic system explicitly, or the parameters depend on time, then the dynamic system will be *nonautonomous*. An N-dimensional nonautonomous dynamic system can be converted into an $(N + 1)$-dimensional autonomous system by the introduction of time t as an additional dependent variable by the phase equation $\mathrm{d}t/\mathrm{d}\tau = 1$, and consider τ the new independent time variable. We shall therefore only consider autonomous systems in the following. If the parameters depend on the state of the dynamic system, then the parameters will give rise to additional nonlinearities.

In order to simplify the nonlinear dynamic problem, the system may be idealized within certain limits. The simplest assumption is that all the parameters are independent of the state of the system, but it is wrong for all real mechanical systems. In many cases, however, it is possible to find intervals of the action in which certain parameters in practice (i.e., within a given small tolerance) remain constant. It may for example happen that the action of a damper in the problem of interest is limited to such a small velocity interval that its nonlinear character may be neglected. Such simplifications in the theoretical model must, however, be made with great caution, because they may change the solutions of the problem *qualitatively* and/or *quantitatively* and important properties of the solution may be lost.

5.3 ON THE GEOMETRY OF NONLINEAR AUTONOMOUS DYNAMIC SYSTEMS WITH CONSTANT PARAMETERS

5.3.1 Some Definitions

The geometric theory of nonlinear dynamic systems provides the best access to an understanding of the behavior of such systems. A dynamic system is a system of first-order ordinary differential equations that often is connected with a system of algebraic or transcendental equations, which define constraints in the system. The dependent variables x_i, $i = 1, 2, \ldots, N$ are called *state variables*, and they are all real. $\boldsymbol{x}$ is the N-dimensional vector $\boldsymbol{x} = \{x_i\}$, $i = 1, 2, \ldots, N$. Time, t, is the independent variable. We denote $\mathrm{d}x_i/\mathrm{d}t$ by $\dot{x}_i$ and correspondingly $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$ by $\dot{\boldsymbol{x}}$. $\boldsymbol{\lambda} = \{\lambda_j\}$, $j = 1, 2, \ldots, M$ denotes a set of M real-valued parameters. A *parameter-dependent nonlinear dynamic system* can then be written in the form

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda}). \tag{5.1}$$

$\boldsymbol{F}$ is an N-dimensional real-valued "sufficiently smooth" nonlinear vector function of the N-dependent variables x_i. By *function* we always mean an expression from which the value of the function is determined *uniquely* by the argument. If uniqueness is not guaranteed, we speak of a *relation*. An example of a relation is the expression $f = \pm\sqrt{x}$, since there exist two real values of f for every positive value of x. When we deal with functions, we call the set of all permissible values of the arguments x_i *the domain of the function* $\boldsymbol{F}$. The set of values that the function can assume for all permissible variations of the arguments is called *the range of the function* $\boldsymbol{F}$.

The number of degrees of freedom (DOF) K of a mechanical system equals the number of generalized coordinates of the system. The mechanical system will normally be modeled mathematically by a system of K ordinary differential equations of second order in the dependent variable q_i. The trivial substitution $q_i = x_{2i-1}$ and $\dot{q}_i = x_{2i}$, $i = 1, 2, \ldots, K$ will however transform the system of K second-order differential equations into a dynamic system of $2K$ first-order differential equations, and we then set N equal to $2K$.

We now introduce a Cartesian coordinate system with the state variables x_i as the coordinates. The system will be N-dimensional and it is denoted *the state space* or *phase space*. We can plot the values $x_i(t)$ that satisfy Equation 5.1 for fixed parameter values λ_j in the phase space. We then obtain a set of curves, depending on the initial conditions, which are called *trajectories* or *integral curves*. They are the geometric state space representations of the solutions of the initial value problems defined by Equation 5.1. t is a curve parameter and associated with each trajectory is a direction showing how the state of the system changes as time increases. The complete picture is called *the phase diagram* or *phase portrait* of the system. A point in the phase space is called an *image point* and the trajectory through an image point is often termed a *phase trajectory* in order to distinguish it from the trajectory of the motion in the physical world, which is $\boldsymbol{q}(t)$. The speed of an image point is called *the phase speed* not to be confused with the physical speed of the elements in the system. In autonomous systems $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda})$ with $\boldsymbol{F}$ sufficiently smooth, *trajectories cannot cross each other*, meaning that the motion of the system on a trajectory is determined *uniquely* by any point on that trajectory. An *equilibrium point* is an image point that does not change with time, so its phase trajectory is just a point. It corresponds to a *stationary state* in the state space and a *stationary solution* of the dynamic system. A point, in which the slope of the trajectory is not defined, is called a *singular point*; all other points are called *regular points*. An equilibrium point is a singular point.

Since a periodic solution will always return to any given state with its period T, *the trajectory of a periodic solution must be a closed curve.*

Dissipative autonomous systems are *contracting in the phase space* in the sense that a small volume bounded by a bundle of trajectories and lying between t_0 and $t_0 + \mathrm{d}t$ will decrease when it moves along its bundle of trajectories with growing t (see Figure 5.2 for two examples).

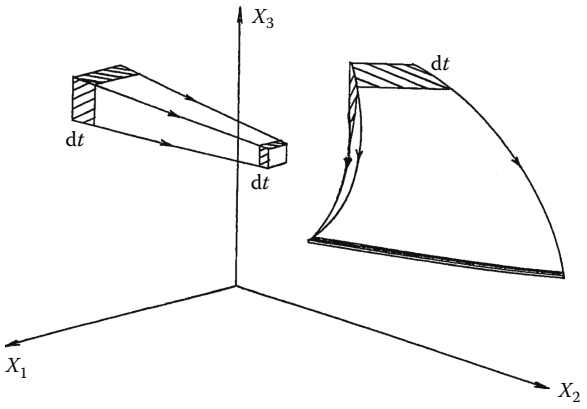


FIGURE 5.2 Examples of trajectories of a dissipative dynamic system in a state space.

5.3.2 Stability Concepts

We now introduce the four most important concepts of stability. First, we define the stability of equilibrium points or stationary solutions. Given

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \lambda) \tag{5.1}$$

with the equilibrium point $\mathbf{0}$. It means that

$$\boldsymbol{F}(\mathbf{0}; \lambda) = \mathbf{0}.$$

A simple coordinate transformation will take any one point in the state space to $\mathbf{0}$.

For equilibrium points, we first define *local stability* or *Lyapunov stability.*

Definition 5.1

If there for all $\varepsilon > 0$ *exist* $\delta > 0$ *(to every* δ *there corresponds an* ε*) such that* $\|\boldsymbol{a}\| < \delta \rightarrow \|\boldsymbol{x}(t)\| < \varepsilon$ *for all t, where* $\boldsymbol{a}$ *denotes an initial condition and* $\| \ \|$ *denotes a suitable norm (or measure) then the point* $\mathbf{0}$ *is* locally stable *or* Lyapunov stable.

The definition states that a trajectory, which starts inside a ball with radius δ and center in $\mathbf{0}$ (the domain needs not be spherical), will never leave the ball with radius ε and center in $\mathbf{0}$ (the domain is not necessarily spherical) (see Figure 5.3).

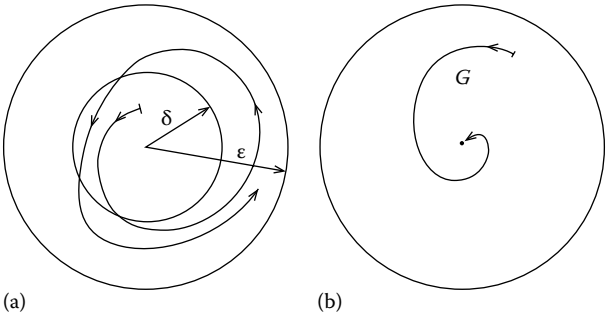


FIGURE 5.3 Trajectories near the stable point $\mathbf{0}$. (a) Locally stable and (b) asymptotically stable.

Definition 5.2

If $\boldsymbol{0}$ *is locally stable for all trajectories that start inside a domain G, which contains* $\boldsymbol{0}$*, and the following relation in addition holds*

$$\|\boldsymbol{x}(t)\| \to 0 \quad \text{for } t \to \infty$$

then $\boldsymbol{0}$ *is* asymptotically stable.

It is the second and most used concept of stability for dissipative systems. The definition states that $\mathbf{0}$ is asymptotically stable if all trajectories that start in G or enter G ultimately will come arbitrarily close to $\mathbf{0}$.

Next, we define *orbital stability*. We now consider an orbit or closed trajectory H and a tube of radius ε around H. Please remember that a closed trajectory in phase space represents a periodic solution in the physical space.

Definition 5.3

If all paths starting from initial points $\boldsymbol{a}^*$ *within a distance* δ *from any point* $\boldsymbol{a}$ *on H remain permanently within a tube of radius* ε *for* $t \to \infty$ *then the orbit H is* orbitally stable.

If all the paths come arbitrarily close to H for $t \to \infty$, then the orbit H is *asymptotically orbitally stable* (see Figure 5.4). This definition can be extended to other steady-state solutions of the dynamic system.

Points or paths that are not stable (locally or orbitally) are *unstable*. They are unstable under *infinitesimally small disturbances*.

All the stability concepts mentioned earlier refer to stability with respect to disturbances of *initial conditions*. The last concept of stability we shall use refers to disturbances of the vector function

$$\boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda})$$

We, therefore, finally define *structural stability*.

Definition 5.4

A dynamic system is structurally stable *if small disturbances of the vector function* $\boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda})$ *lead to equivalent phase diagrams*.

An observable phenomenon should always be modeled by a structurally stable system, because small errors in the modeling and fluctuations in the physical system are unavoidable. A conservative dynamic system is structurally unstable according to this definition, since the addition of an ever so small dissipation will change the phase diagram qualitatively.

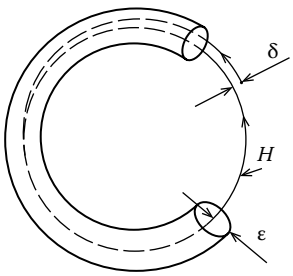


FIGURE 5.4 An asymptotically orbitally stable trajectory.

For the investigations of asymptotic stability of equilibrium points, *Lyapunov's first method* may be applied. With no loss of generality, we shall again assume that the equilibrium point is **0**. Given the autonomous system:

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda}) \tag{5.1}$$

with $\boldsymbol{F}(\boldsymbol{0}; \boldsymbol{\lambda}) = \boldsymbol{0}$

We assume that *the Jacobian*

$$\boldsymbol{J} = \left[\frac{\partial f_i}{\partial x_j}\right], \quad i, j = 1, 2, \ldots, N,$$

exists and provides the matrix $\boldsymbol{J_0}$ in $\boldsymbol{x} = \boldsymbol{0}$. We then write Equation 5.1 in the form:

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda}) = \boldsymbol{J_0 x} + \boldsymbol{G}(\boldsymbol{x}; \boldsymbol{\lambda}), \quad \boldsymbol{J_0} = \boldsymbol{J_0}(\boldsymbol{\lambda}).$$

The equation

$$\dot{\boldsymbol{x}} = \boldsymbol{J_0 x} \tag{5.2}$$

is Equation 5.1 *linearized around* **0**, and $\boldsymbol{G}(\boldsymbol{x}; \boldsymbol{\lambda})$ is a remainder, which contains all the nonlinear contributions.

Theorem 5.1

Let **0** be an equilibrium point of $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda})$.

1. *If the Jacobian* $\boldsymbol{J_0}$ *has neither zero nor purely imaginary eigenvalues*
2. $\|\boldsymbol{G}(\boldsymbol{x}; \boldsymbol{\lambda})\|/\|\boldsymbol{x}\| \to 0$ for $\|\boldsymbol{x}\| \to 0$

then it is a necessary and sufficient condition for asymptotic stability *of* **0** *that all solutions of* $\dot{\boldsymbol{x}} = \boldsymbol{J_0 x}$ *are* locally stable.

In other words, under the conditions stated, the *linearized system* determines the stability of the full nonlinear system in a certain neighborhood of **0**. Vehicle dynamic systems are dissipative systems, and local stability of $\dot{\boldsymbol{x}} = \boldsymbol{J_0 x}$ is therefore guaranteed, when all the eigenvalues of $\boldsymbol{J_0}$ have negative real parts. In Definition 5.1, δ may be very small, and in Theorem 5.1 the condition 2 may only be satisfied in a very small neighborhood of **0**. Therefore, the stability is local in the sense that the stability of the trivial solution, $\boldsymbol{x} = \boldsymbol{0}$, is only guaranteed for infinitesimally small disturbances. The same limitation holds for linear stability analyses.

In order to linearize a dynamic system, *it must be determined by a function* in order to guarantee uniqueness of the point you linearize around. When the function is expanded in a series around the point of expansion, the linear part must be the lowest order contribution when the variables approach the point of expansion; otherwise, the nonlinear terms will dominate the behavior of the function around the point of expansion.

Therefore, the coefficients of the linear terms in the expansion *must be finite and nonzero.* Therefore, $\boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda})$ in (5.1) cannot be linearized around **0** when for instance $\boldsymbol{F}(\boldsymbol{x}; \boldsymbol{\lambda}) = \pm\sqrt{x}$, x^2, or $\cos x$.

5.3.3 Equilibrium Solutions

We now consider equilibrium points and cycles of dissipative, autonomous problems with the parameters fixed:

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) \tag{5.3}$$

5.3.3.1 Equilibrium Points

We shall only consider equilibrium points for which Theorem 5.1 holds. The equilibrium points and their stability are then characterized by the sign of the eigenvalues of the Jacobian if the eigenvalues are real, and by the sign of the real part of the eigenvalues if the eigenvalues are complex. Since we have assumed that the state variables as well as the vector function $\boldsymbol{F}(\boldsymbol{x})$ are real, then the complex eigenvalues will always appear as complex conjugate pairs. If all the eigenvalues are real and of the same sign, then the equilibrium point is called a *nodal point* or a *node*. If all the eigenvalues are negative, then the nodal point is asymptotically stable; otherwise, it is unstable. If some eigenvalues are complex then the equilibrium point is called a *spiral point* or a *spiral* or a *focus*, and if the real parts of all the eigenvalues are negative then the spiral is stable. Notice that the corresponding eigenvectors span directions—in the case of nodal behavior—or planes in the case of spiral behavior (see Figure 5.5). In state spaces of dimension three or higher, an equilibrium point may have nodal properties along some directions and spiral properties in some surfaces.

If the real part of the eigenvalues has different signs, then the equilibrium point is a *saddle point* or just a *saddle*. A saddle i*s always unstable* since the eigenvectors belonging to the eigenvalues with positive real parts are directions of repulsion (see Figure 5.6). Notice that an equilibrium point is itself a trajectory, which consists of just that point. The trajectories that approach a stable equilibrium point will only get arbitrarily close to it in any finite time and never reach it. The theorem that trajectories in an autonomous system cannot cross each other thus still holds true.

The set of all the points on the trajectories that approach a stable equilibrium point is called *the basin of attraction* of that particular equilibrium point. When several stable equilibrium points coexist, it is particularly important to determine the basins of attraction for each of them, since one or more of the equilibrium points may be undesirable states (see Figure 5.7). Initial values that lead to the undesirable states should therefore be avoided.

5.3.3.2 Cycles

Cycles are state space representations of periodic motions. They are closed curves in the state space because of the periodicity. Before we consider cycles in more detail, it will be convenient to introduce the concept of a *discrete map* or just *map*. It is the discrete equivalent of our dynamic system (Equation 5.1).

An equation in the form:

$$\boldsymbol{y}_{n+1} = \boldsymbol{F}(\boldsymbol{y}_n; \boldsymbol{\lambda}), \quad n = 1, 2, \ldots$$

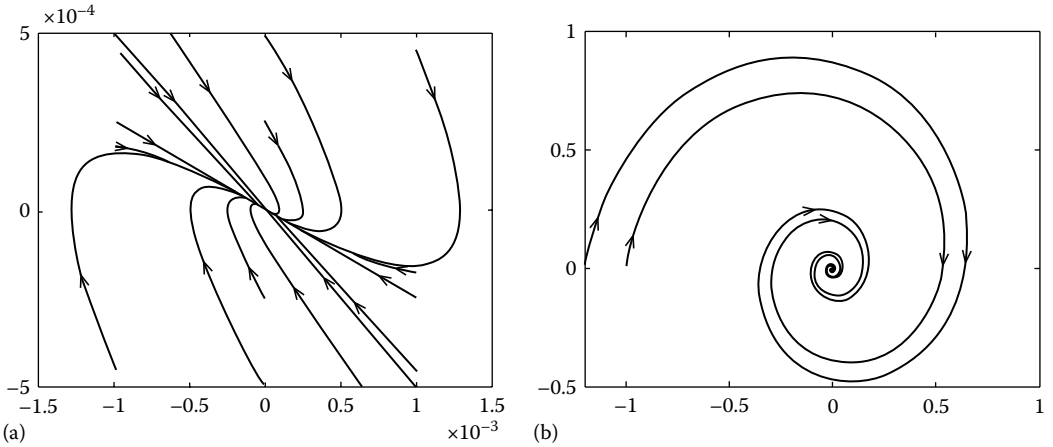


FIGURE 5.5 A stable node (a) and a stable spiral (b). If the directions of motion on the trajectories are reversed, then the equilibrium points are unstable.

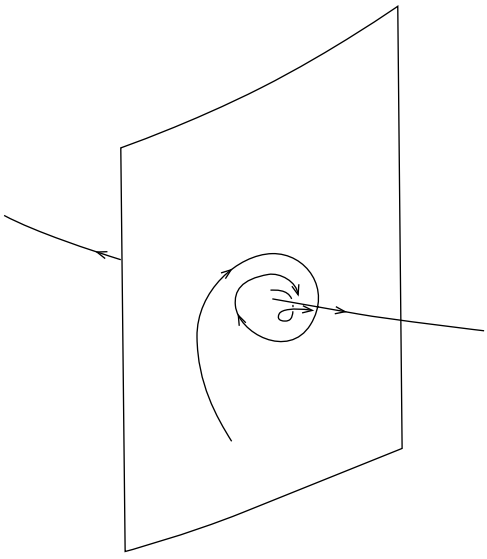


FIGURE 5.6 An example of a saddle in a 3D phase space. All saddles are always unstable. (A saddle in a phase plane is shown in Figure 5.7.)

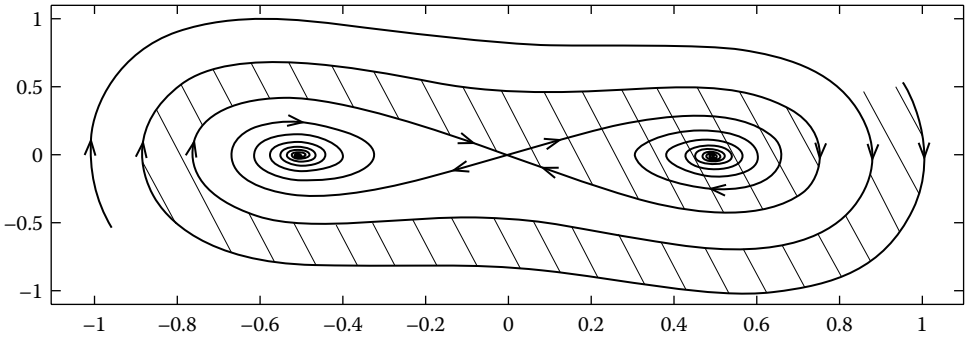


FIGURE 5.7 Two stable spirals separated by a saddle at $\boldsymbol{x} = \mathbf{0}$. The "incoming" asymptotic trajectories (insets) split the domains of attraction of the two spirals. (Notice how a saddle appears in a phase plane—compare with Figure 5.6.)

maps the nth vector $\mathbf{y}_n$ into the $(n+1)$st vector $\mathbf{y}_{n+1}$ by the *mapping function* $\boldsymbol{F}$. $\boldsymbol{F}$ is an N-dimensional real-valued "sufficiently smooth" nonlinear vector function of the N-dependent variables

$$y_n^i,\ i = 1,\ 2,\ \ldots,\ N.$$

When we investigate the dynamics of periodic or other recurrent motions, we shall use *the Poincaré map*. We introduce a plane in the state space and orient it in such a way that the trajectory intersects it transversally, that is, it is not a tangent plane to the trajectory at any point (Figure 5.8). Such a plane is called a *Poincaré section*. A cycle will intersect the Poincaré section at least twice but we now choose to consider only one of the intersection points $\boldsymbol{y}_0$. We now follow the cycle from $\boldsymbol{y}_0$ all the way around the trajectory until it intersects the Poincaré section again *in the same direction*. We call that point $\boldsymbol{y}_1$. It is the image of $\boldsymbol{y}_0$ *under the Poincaré map* $\boldsymbol{P}$. $\boldsymbol{P}$ thus takes the point $\boldsymbol{y}_0$ into $\boldsymbol{y}_1$, and in the case of a cycle $\boldsymbol{y}_1$ will in general be identical to $\boldsymbol{y}_0$ so $\boldsymbol{P}$ reduces to the identity map $\boldsymbol{I}$. The domain and range of the map $\boldsymbol{P}$ lie in the Poincaré section only. You can think of the Poincaré map as a stroboscopic image of a point moving on a recurrent trajectory.

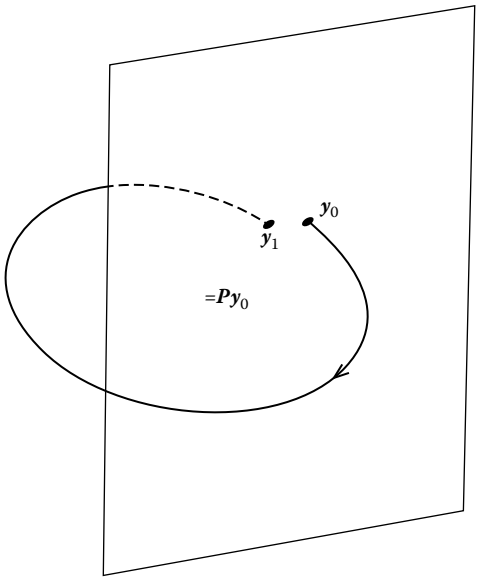


FIGURE 5.8 The Poincaré map.

The Poincaré map transforms a periodic solution of Equation 5.1 into a stationary solution in a Poincaré section of

$$\boldsymbol{y}_{n+1} = \boldsymbol{P}(\boldsymbol{y}_n; \lambda). \tag{5.4}$$

An equilibrium or fixed point $\boldsymbol{y}_0$ of the Poincaré map $\boldsymbol{P}$ has the property:

$$\boldsymbol{y}_0 = \boldsymbol{P}^m(\boldsymbol{y}_0; \lambda), \quad m = 1, 2, \ldots.$$

where $\boldsymbol{P}^m$ denotes the mth iterate of the Poincaré map $\boldsymbol{P}$. We can now apply the theory of stability of solutions of discrete maps to the equilibrium point for $\boldsymbol{P}$. In analogy with the theory for continuous systems, a linearization $\boldsymbol{P}_y$ of $\boldsymbol{P}$ is calculated and evaluated at the equilibrium point. The corresponding linear map will then be

$$\boldsymbol{y}_{n+1} = \boldsymbol{C}\boldsymbol{y}_n \tag{5.5}$$

where the elements of the matrix $\boldsymbol{C}$ are the elements of $\boldsymbol{P}_y$ evaluated at the equilibrium point. We now introduce the linear transformation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}$$

in Equation 5.5 and obtain

$$\boldsymbol{A}\boldsymbol{z}_{k+1} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{z}_k$$

We assume that $\boldsymbol{A}$ is non-singular and multiply from the left with $\boldsymbol{A}^{-1}$ and obtain

$$\boldsymbol{z}_{k+1} = \boldsymbol{M}\boldsymbol{z}_k, \quad \boldsymbol{M} = \boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{A} \tag{5.6}$$

We choose $\boldsymbol{A}$ so that $\boldsymbol{M}$ has a Jordan canonical form. If the eigenvalues ρ_m of $\boldsymbol{C}$ are distinct, $\boldsymbol{M}$ is a diagonal matrix with entries $\rho_1, \rho_2, \rho_3, \ldots, \rho_N$. Then the kth iterate of Equation 5.6 can be rewritten as

$$z_{k+1}^{(m)} = \rho_m z_k^{(m)}, \quad m = 1, 2, \ldots, N$$

where $z^{(m)}$ is the mth component of z. It follows from Equation 5.6 that for $k \to \infty$,

$$z_k^{(m)} \to 0 \quad \text{if } |\rho_m| < 1$$

$$z_k^{(m)} \to \infty \quad \text{if } |\rho_m| > 1$$

$$z_k^{(m)} = z_0^{(m)} \quad \text{if } \rho_m = 1$$

$$z_{2k+1}^{(m)} = -z_0^{(m)} \quad \text{and}$$

$$z_{2k}^{(m)} = z_0^{(m)} \quad \text{if } \rho_m = -1.$$

We see that the location of the eigenvalues ρ_m of $\boldsymbol{C}$ in the complex plane determines the stability of the equilibrium point $\boldsymbol{y}_0$. If all the eigenvalues of $\boldsymbol{C}$ lie inside the unit circle in the complex plane, the equilibrium point $\boldsymbol{y}_0$ is asymptotically stable. If at least one of the eigenvalues lies outside the unit circle, then $\boldsymbol{y}_0$ is unstable. If at least one of the eigenvalues lies on the unit circle and all the rest inside, then a linear stability analysis will be inconclusive.

If $0 < \rho_m < 1$ for all ρ_m disturbances will decay monotonously, and if for at least one eigenvalue ρ_k $-1 < \rho_k < 0$ disturbances will decay in an oscillatory fashion. We call this behavior *flipping*. If at least one eigenvalue is complex, then the disturbance will rotate around the equilibrium point while it decays or grows depending on whether all eigenvalues are inside the unit circle in the complex plane or not. A cycle is thus *orbitally asymptotically stable*, if all the eigenvalues of $\boldsymbol{C}$ lie inside the unit circle in the complex plane [4,6].

As before, the set of all the points on the trajectories in the state space that approach a stable cycle is called *the basin of attraction* of that particular cycle. A stable steady solution is called an *attractor*. It can be a point, a cycle, or a higher-dimensional or even fractal-dimensional structure. An unstable stationary periodic solution is most often a *saddle cycle*. When several attractors coexist, it is desirable to determine the basins of attraction for each of them in order to guard against unpleasant surprises caused by initial conditions that will lead to one of the undesirable attractors.

All that have been said earlier about the calculation of stability of cycles only serve to help the reader understand the concept and the relation between cycles in state space and Poincaré maps. It is not a recipe for a stability analysis of periodic solutions of nonlinear problems. The reason is that *the author has never seen any nonlinear dynamic system*

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \lambda)$$

with a periodic solution, where the corresponding Poincaré map

$$\boldsymbol{y}_{n+1} = \boldsymbol{P}(\boldsymbol{y}_n; \lambda), \quad n = 1, 2, \ldots$$

could be determined explicitly. Instead *Floquet theory* (see [4,6]) is often applied directly to the continuous problem (Equation 5.1). We shall however here present an easier method that is suitable for a numerical analysis of a known periodic solution of a dynamic problem.

Given a periodic solution to Equation 5.1, $\boldsymbol{x}_0(t, T_0)$ with the known period T_0. Choose a value of t, say t_0 and calculate $\boldsymbol{x}_0(t_0, T_0)$. Put a plane S that contains $\boldsymbol{x}_0(t_0, T_0)$ in the state space transversal to the trajectory of $\boldsymbol{x}_0(t, T_0)$. Add a small disturbance $\boldsymbol{y}$ to $\boldsymbol{x}_0(t_0, T_0)$ in S, so both $\boldsymbol{x}_0(t_0, T_0)$ and $\boldsymbol{x}_0(t_0, T_0) + \boldsymbol{y}$ lie in S. Calculate the distance $d_0 = \|\boldsymbol{y}\|$ between $\boldsymbol{x}_0(t_0, T_0)$ and $\boldsymbol{x}_0(t_0, T_0) + \boldsymbol{y}$. Integrate the dynamic problem (Equation 5.1) with the initial condition $\boldsymbol{x}_0 + \boldsymbol{y}$ for one or a couple of periods T_0 and obtain another point $\boldsymbol{x}_1 = \boldsymbol{x}_0(t_0, T_0) + z$ in H. Calculate the distance $d_1 = \|z\|$ between $\boldsymbol{x}_0(t_0, T_0)$ and $\boldsymbol{x}_0(t_0, T_0) + z$. Finally, calculate the ratio $\rho = d_1/d_0$. If $\rho < 1$, then the periodic solution $\boldsymbol{x}_0(t, T_0)$ is *asymptotically orbitally stable*. Otherwise, it is unstable.

In conclusion, we shall note that recurrent solutions that are characterized by several periods are also possible in nonlinear dynamic systems. They are called *multi-periodic* solutions. In case all the periods are commensurate—that is, the ratio between each pair of periods is a rational number—then the resulting motion is periodic. If that is not the case, then the resulting motion is *quasi-periodic*.

5.3.3.3 Chaos

It is a well-known fact that deterministic nonlinear dynamic systems can behave erratically. We want here to make a clear distinction from random motion. We shall call a process *random*, when uncertainties in the dynamic system appear as probabilistic relations or state variables. Such systems are not considered here.

The kind of erratic behavior in deterministic systems is therefore of another character, and it is often called *chaos*. There exist both *chaotic attractors* also called *strange attractors* and *chaotic repellers*. Since only the attractors can be observed in real-life situations, we shall only describe those. There is a third important form of chaotic motion—*chaotic transients*—which will be discussed briefly later.

Chaotic motion is *recurrent* but it is neither periodic nor quasi-periodic. Since it is recurrent but nonperiodic, the trajectories of a chaotic motion will "return" to the neighborhood of points they passed earlier without ever closing up at any point. Furthermore, all trajectories on a chaotic attractor are orbitally unstable. They reject each other like saddle cycles, so in order to hold the bundle of all trajectories inside a bounded domain a *reinjection mechanism* must be at work in the dynamic process. That reinjection mechanism takes a bundle of trajectories and folds them back onto the other trajectories without touching any of them. This action is the cause of *mixing* in chaotic dynamics. In a 3D phase space, the attractor therefore consists of many infinitely thin "layers" in which the trajectories are imbedded—a structure that can be described by the mathematical concept of a *fractal dimension*. The layers are separated and so close to each other that the attractor seems 2D in the 3D phase space.

In Figure 5.2 (right-hand side), an example of a set of trajectories that repel each other in a surface and contract in the direction normal to the surface is shown. This action is typical for chaos. Due to the dissipation, the volume elements in phase space must decrease with time—as earlier mentioned—but that property does not prohibit a development as shown in Figure 5.2. We only need still to fold the two diverging "wings" back on top of each other and bend the whole structure around a cylinder in order to obtain a picture of the structure of a simple chaotic attractor (see Figure 5.9). It shows the simplest known chaotic attractor—the so-called *Rössler attractor.* Because of the folding action, chaos in autonomous systems only exists in three- or higher-dimensional systems.

$$\dot{x} = -y - z; \quad \dot{y} = x + 0.398y; \quad \dot{z} = 2 + z(x - 4)$$

A chaotic attractor can, according to Thompson and Stewart [1], briefly be described as "a stable structure of long-term trajectories in a bounded region of phase space, which folds the bundle of trajectories back onto themselves, resulting in mixing and divergence of nearby states."

When chaos develops in a nonlinear dynamic system—in a real-life situation or a numerical simulation—the very first indication is a seemingly erratically oscillating behavior of one or more components of the system. A hallmark of chaotic motion is its *sensitivity to changes*

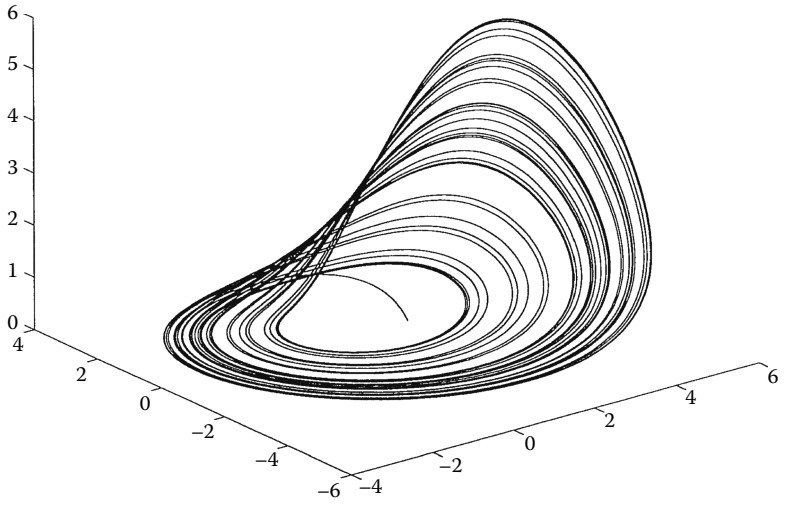


FIGURE 5.9 Rössler's chaotic attractor. It is a steady solution of the dynamic system.

in the initial conditions, but there exist other indicators. According to Moon [2], the following features characterize chaos:

1. Broad spectrum of Fourier transform when the motion is generated by a single frequency (measured by fast Fourier transform [FFT]).
2. Fractal properties of the motion in phase space, which denote a chaotic attractor (measured by Poincaré maps or fractal dimensions).
3. The complexity of the motion (measured by Lyapunov exponents). See a book on nonlinear dynamics [1–9].

When one suspects a motion to be chaotic, the first test is to watch a time series of a component of the motion. Then look at a phase diagram and at least one Poincaré section (see Figure 5.10). Next—if possible—repeat the process with slightly changed initial conditions. If the values of the state variables are changed in regions of the long-time evolution when the initial conditions are changed, then the process is chaotic. You can also quantify the degree of disorder by computations of measures of the fractal dimension of the attractor or—the author's own favorite—a calculation of the largest Lyapunov exponent(s). If a series of calculations of the largest Lyapunov exponent converges to a positive value, then the motion is chaotic.

When the dynamics is chaotic and multiple attractors exist, then the boundaries of the basins of attraction or the basins themselves may be *fractal dimensional*. As a consequence thereof, there exist points $\boldsymbol{X}_n$ in one of the basins of attraction with the property that there exists at least one point in the other basin of attraction in any neighborhood of $\boldsymbol{X}_n$. In other words, when the boundary of a domain of attraction or the domain itself is fractal dimensional then the result of an experimental investigation of the dynamic problem is in general unpredictable.

We shall conclude this brief introduction to chaos by defining *transient chaos*. It is a motion that looks chaotic and appears to show the characteristic features of a chaotic attractor, but it eventually settles into a regular motion or a chaotic attractor. It often appears when multiple attractors exist in phase space. Dynamic systems with chaotic transients must be treated with caution. The transients often decay with ever-decreasing oscillations into a regular motion, but cases have been found where chaotic transients have amplitudes, which are larger than the amplitudes of the attractor, which the transient approaches. Such chaotic transients can be the cause of catastrophic developments in a dynamic system. Pascal [10] has investigated the derailment of a two-axle freight wagon, which moved chaotically.

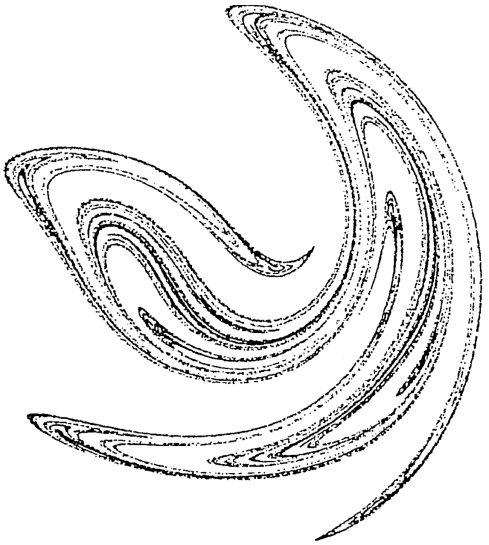


FIGURE 5.10 A Poincaré section plot of one of Ueda's chaotic attractors. The banded structure is typical of chaos. (Thompson, J.M.T. and Stewart, H.B., *Nonlinear Dynamics and Chaos*. 1986. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. With permission.)

In general, chaos is considered the ultimate development of the dynamics of a parameter-dependent system when one or several parameters grow. However, in chaotic systems patterns may develop whereby a reorganization of the dynamic system in clear structures occurs. Maicz et al. [11] treat an example from railway vehicle dynamics.

5.4 PARAMETER-DEPENDENT SYSTEMS AND BIFURCATIONS

We shall concentrate the survey on systems with only one varying parameter λ. It is called a *controlled* parameter. In vehicle dynamic problems the speed of the vehicle, V, is such a parameter. The solutions $\boldsymbol{x}(t; \lambda)$ of the nonlinear dynamic problem

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; \lambda) \tag{5.1}$$

will then depend on the parameter λ as well as on time. When we vary λ in a continuous fashion, the solution will change continuously (according to our assumptions) but what is more important, *the number of solutions may change*, that is, in addition to the already known solution other solutions may appear or disappear at certain parameter values. Obviously our dynamic system is not structurally stable at such parameter values, since an infinitesimal change in the parameter value will change the phase diagram. These features will be depicted in a *parameter-state space diagram*, where the parameter λ is the independent variable. In order to be able to present the diagram in a plane, the states in the phase space are represented by a norm for instance the maximum amplitude or a Euclidean norm.

Very often in nonlinear dynamics, a change in the number of solutions happens through a *branching* or *bifurcation* from an existing solution under a continuous change of the parameter. Another common form of bifurcation is a solution that "turns around" in the sense that the sign of the parameter increments changes in order to follow the solution in the parameter-phase space diagram. Such a bifurcation is called a *fold* or *saddle-node bifurcation*, because the solution changes stability when "it moves around the fold" (see Figure 5.11). We call all these changes *local bifurcations*. They are very important for the analysis of the dynamics of the system, and they occur only in nonlinear systems.

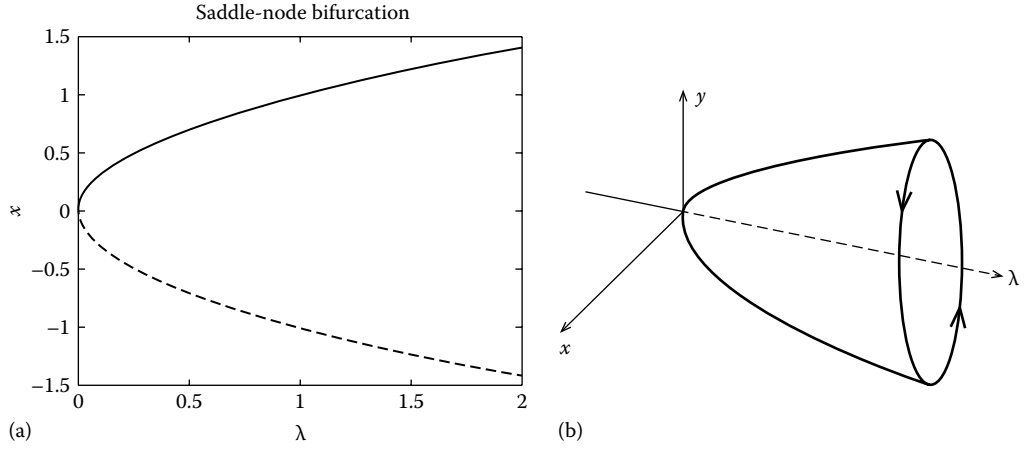


FIGURE 5.11 A saddle-node bifurcation (a) and a Hopf bifurcation (b). A full line denotes a stable and a broken line an unstable solution.

They have in common that the parameter values where bifurcation occurs—*the bifurcation points*—can be found by an analysis of the parameter-dependent eigenvalues of the associated Jacobian. We shall assume in the following that the solution we want to investigate—the original solution—is *asymptotically stable* to start with. In all the cases, the original solution will *lose stability* in the bifurcation point. This is the connection to the well-known linear stability theory.

We can also introduce structural stability of bifurcations. The phrase, however, is hardly ever used in that context. Instead, we denote a bifurcation *generic*, that is, "commonly occurring" when a small change of the dynamic system does not change the properties of the bifurcation.

There are only two generic bifurcations from an equilibrium point under one control parameter. They are the earlier mentioned *saddle-node*, when the largest eigenvalue of the Jacobian becomes positive, and the *Hopf bifurcation* (named after the mathematician, who first described it), which is a bifurcation of a *periodic solution* (see Figure 5.11). In that case, a complex conjugate pair of eigenvalues of the Jacobian crosses the imaginary axis in the complex plane into the positive real half plane. These two kinds of bifurcations are very common in vehicle dynamics.

If we allow more simultaneously acting control parameters, then two additional generic bifurcations are possible. They are the *transcritical bifurcation* and the *pitchfork bifurcation* or *symmetry breaking bifurcation*, when it occurs in systems that are symmetric in some way (see Figure 5.12).

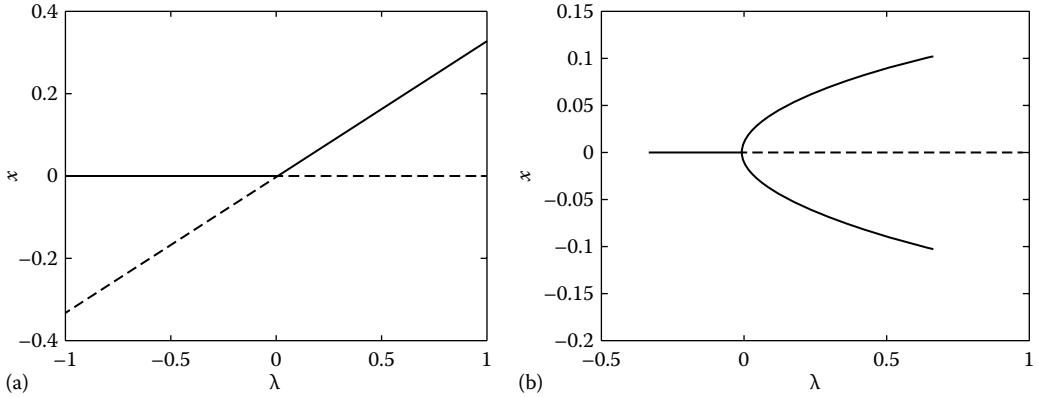


FIGURE 5.12 A transcritical bifurcation (a) and a pitchfork bifurcation (b). A full line denotes a stable and a broken line an unstable solution.

In both cases the largest eigenvalue of the Jacobian becomes positive when λ grows through the bifurcation point, and a computation of the *normal form* (see [6]) is necessary to distinguish them from the saddle-node bifurcation. Today, *center manifold reduction* and *normal forms* are commonly used for local bifurcation analyses. The center manifold reduction theorem states that the change of dynamic behavior of an *N*-dimensional dynamic system around the generic bifurcations can be described by the change of dynamic behavior on a curve or 2D surface in the parameter-phase space. The curve or surface is called a *center manifold* (see Section 5.5). This important theorem guarantees that the bifurcations characterized by one or two eigenvalues of the Jacobian crossing the imaginary axis in the complex plane are the only generic ones that will be found in any *N*-dimensional dissipative dynamic system. The normal form is the simplest possible expression that is necessary to characterize the nonlinear dependence of the amplitudes of the solutions on the parameter near the bifurcation point.

All these bifurcations are also generic for cycles. Only instead of an eigenvalue crossing the imaginary axis, an eigenvalue must escape from the unit disk. A Hopf bifurcation from a cycle creates generically a quasi-periodic motion. Its trajectory lies on a torus in state space and the bifurcation is often called a *Naimark bifurcation*. For cycles there exists an additional bifurcation, the *flip bifurcation* or *period-doubling bifurcation*. In a flip bifurcation, a new periodic motion develops with twice the period of the original motion. A plot of the trajectory of a period-two solution could be the edge of a Möbius strip. The best-known transition to chaos is an infinite sequence of period-doubling bifurcations.

For cycles the analog of "the eigenvalue of the Jacobian crossing into the positive real half plane" is "the Floquet multiplier crossing the unit circle in the complex plane." The Floquet multiplier in dynamic systems is the analog of ρ_m introduced in Section 5.3.3.2. If the Floquet multiplier crosses "1," then we have a fold, or a transcritical, or a pitchfork bifurcation, where again additional calculations are necessary to determine the type. If the Floquet multiplier crosses "–1" we have a period-doubling bifurcation, and if it crosses almost anywhere else (a few special points are excluded) it is a Naimark bifurcation.

There exist conditions for the stability of the bifurcating solution—or new branch—from equilibrium points. For the case of a Hopf bifurcation, see Figure 5.13.

As a main rule, an analysis of a bifurcation consists of three parts:

1. Determination of all the branches
2. Determination of the stability of the original solution on the other side of the bifurcation point
3. Determination of the stability of the new solution(s)

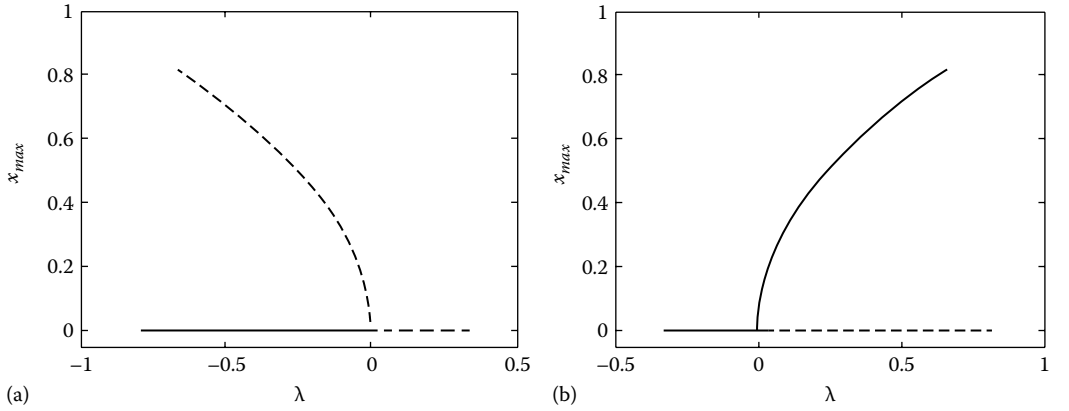


FIGURE 5.13 A subcritical Hopf bifurcation, the periodic solution is unstable (a) and a supercritical Hopf bifurcation, the periodic solution is stable (b). A full line denotes a stable and a broken line an unstable solution.

However, there are other situations in which solutions can appear or disappear. Consider a case where we follow a certain solution in parameter-state space and a parameter value is reached at which a saddle-node appears somewhere in the parameter-phase space away from the solution we follow. Relative to our "certain solution," the saddle-node bifurcation is a *global bifurcation*. There exist many other kinds of global bifurcations, which due to their nature cannot be investigated by "local methods" such as a linearization around a known point in parameter-phase space. For further information, the reader is referred to one of the several books on nonlinear dynamics [1–9].

For the sake of completeness, we just mention *tori* as examples of other equilibrium sets in the phase space. Quasi-periodic functions and multi-periodic functions all lie on tori in appropriate phase spaces. We refer to the literature on nonlinear dynamics for information on their properties and the bifurcations that are connected with them.

5.5 CENTER MANIFOLD REDUCTION THEOREM

The center manifold reduction theorem is the mathematical basis for linking the results we presented in the preceding section to vehicle dynamic systems with their in general high number of DOF. The theorem basically states that all the bifurcations that may be found in high-dimensional parameter-state spaces in general are those described in Section 5.4. The theorem is, however, proven only for systems of differential equations and is valid only for bifurcations from equilibrium points and limit cycles. Furthermore, the dynamic systems must also satisfy certain conditions of smoothness, and *vehicle dynamic systems will rarely satisfy either one of these conditions*. Vehicle dynamic systems are often formulated as differential-algebraic systems, and *the systems are rarely smooth*. In Section 5.7, we shall therefore briefly describe some of the effects discontinuities like impacts and dry friction with stick-slip may have on the bifurcations.

5.6 ON STOCHASTIC DYNAMICS AND THE INFLUENCE OF NOISE

Stochastic elements in the vehicle dynamic models and data represent uncontrolled environmental impacts, thermal fluctuations, or hidden internal dynamics. In sharp contrast to linear systems or compared with systems close to equilibrium, noise and fluctuations in nonlinear dynamics may have a more crucial influence on the behavior of the systems as a result of the rare but large events leading for example to the escape from locally stable states as is found in bistable or multi-stable parameter regions in the vehicle dynamic problems. A channeling action of intrinsic nonlinearities may also influence the systems. Both effects give rise to signal and order amplification by noise or to the emergence of directed motion due to the presence of nonequilibrium fluctuations as discussed in many models and applications. It can be used constructively, in a controlled manner, to enhance nonlinear features leading to ordered behavior or to amplification of signals or deterministic responses. The traditional view on noise was guided by investigations of the linear dynamics of systems close to equilibrium. In such cases, noise introduces disorder and destroys temporal and spatial patterns such as for example symmetries in the system. In most applications, therefore, there have been attempts to minimize the effects of noise.

In a mathematical model of a vehicle dynamic system, noise can be modeled by added perturbations with certain spectral properties or by a representation of driving forces and/or parameters by probability distributions. The parameter dependence of the solutions will be then be expressed by probability distributions. It happens that unknown parameters are a part of the dynamic problem. In order to estimate them in the state space problem, the unknown parameters are often included in the state variables and special methods must be applied for the state estimation. The resulting state space model is in general a nonlinear one, and the extended Kalman filter is then most often used for the state estimation. These cases all lead to a system of *stochastic*

differential equations. For the solution of stochastic differential equations, the reader is referred to the literature (see Øksendal [12]).

For general information on the topic, the following references may prove valuable: The European Science Foundation (ESF) has funded a program on "Stochastic Dynamics: Fundamentals and Applications" (STOCHDYN). It ended in May 2008. Information may be found on the web address: www.esf.org/index.php?id=2266.

When the dynamic systems that we considered in the preceding sections are subjected to noise, some bifurcations may change character. The two generic bifurcations, saddle-node and Hopf, are preserved but the curves that characterize the Hopf bifurcation are smeared out in such a way that the transition to the periodic state takes place in a neighborhood of the bifurcation point of the corresponding deterministic system. An example is shown in Figure 5.14. The bifurcation diagram in Figure 5.14 actually demonstrates the influence of dry friction lateral damping on the dynamics of a railway wheelset (see [13]). Figure 5.14 compares the bifurcation diagram of a wheelset with friction damping with the bifurcation diagram of the same model but with no friction damping. In general, noise will smear out and eliminate a supercritical Hopf bifurcation in the same way as that shown in Figure 5.14.

Under added noise a supercritical pitchfork bifurcation breaks up into a branch with no bifurcation point together with a saddle-node bifurcation in the neighborhood of the bifurcation point of the corresponding system without noise. Near that neighborhood, the system may flip between the resulting two stable branches depending on the noise level (see Figure 5.15). A transcritical bifurcation breaks up under added noise in two different ways. One is into a stable branch and an unstable branch, both without bifurcation points. It means that the bifurcation disappears (see Figure 5.16). Near the bifurcation point of the corresponding noise-free system, the stable system may drift away from the branch if the noise level is sufficiently high. The second possibility is a breakup into two saddle-node bifurcations in the neighborhood of the bifurcation point of the corresponding

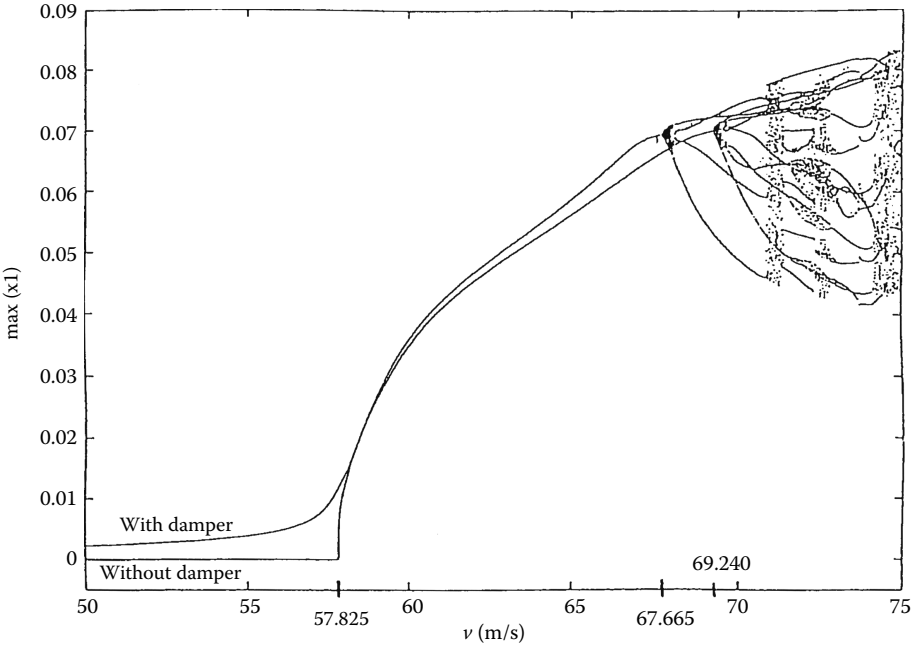


FIGURE 5.14 Two diagrams of the bifurcation in a wheelset model without damper and one with non-smooth dry friction damping.

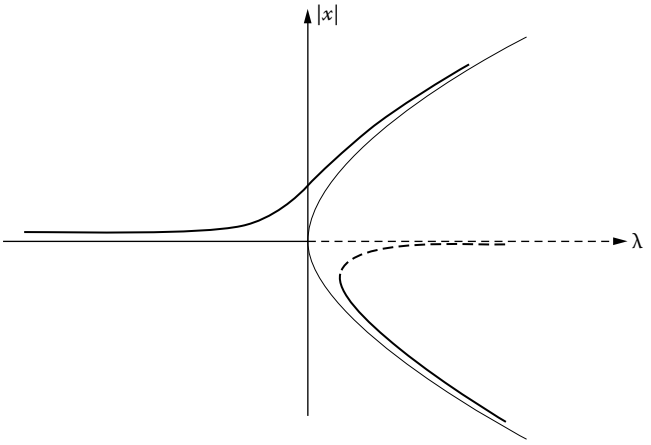


FIGURE 5.15 An example of the breakup of a pitchfork bifurcation under the influence of noise. A full line indicates a stable and a broken line an unstable solution. The thin lines and curves indicate the original pitchfork bifurcation, located in the origin of the parameter-phase diagram.

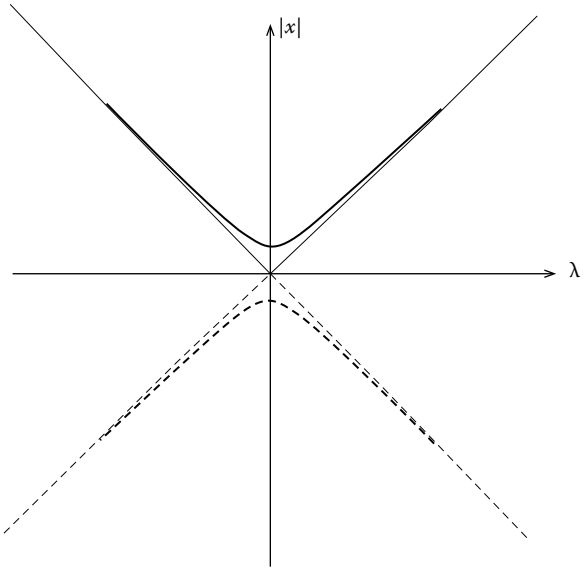


FIGURE 5.16 A first example of the breakup of a transcritical bifurcation under the influence of noise. A full line indicates a stable and a broken line an unstable solution. The thin lines indicate the original transcritical bifurcation, located in the origin of the parameter-phase diagram.

noise-free system (see Figure 5.17). Note that in this case the noise has changed the original transcritical bifurcation from a *continuous* bifurcation to a *dangerous* bifurcation. In the neighborhood of the bifurcation point of the corresponding noise-free system, the post-bifurcation response jumps to a remote attractor, which may be bounded or unbounded. Unbounded responses have disastrous consequences. Typically a bounded response will remain on the path of the new attractor, when the direction of change of the control parameter is reversed resulting in hysteresis.

In stochastic dynamic systems, an additional kind of bifurcation may occur. It is called *stochastic resonance*. It often occurs in bistable systems with noise. When the noise level is low,

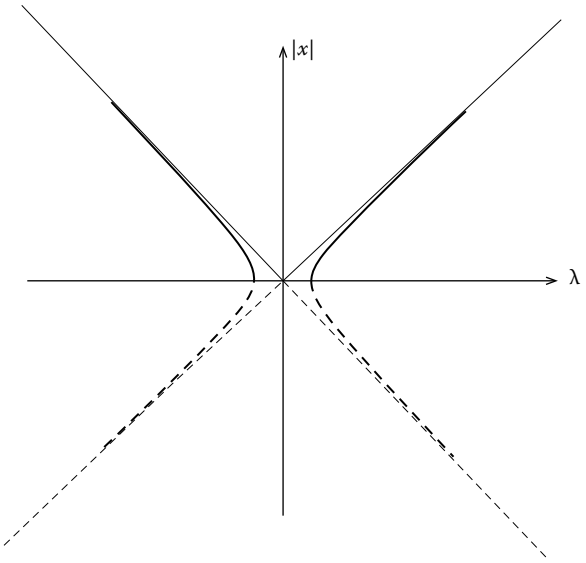


FIGURE 5.17 A second example of the breakup of a transcritical bifurcation under the influence of noise. A full line indicates a stable and a broken line an unstable solution. The thin lines indicate the original transcritical bifurcation, located in the origin of the parameter-phase diagram.

the trajectories of the motion remain within one of the basins of attraction. If a sufficiently large instantaneous singular excitation occurs, then the trajectory will jump from one basin of attraction to another, and then remain there. For large noise intensities, the dynamics is dominated by the noise—an undesirable situation in vehicle dynamic applications. There may, however, exist moderate noise intensities that force the dynamic system to flip between the basins of attraction without dominating the dynamics of the system. Stochastic resonance thus occurs in bistable dynamic systems, when the system response is driven by a combination of a periodic and a wide band stochastic force (noise). The combination of the two forces may then cooperate/compete to make the system switch between the basins of attraction. See an example from Gammaitoni et al. [14] in Figure 5.18.

When the noise is small very few switches occur, mainly at random with no significant periodicity in the system response. When the noise is very strong, a large number of switches occur for each period of the periodic motion and the system response does not show a characteristic periodicity. There exists, however, an optimal value of the noise that corresponds with the periodic motion in such a way that the dynamic system makes almost one switch per period. Such a favorable condition is determined by the matching of two time scales: the period of the periodic motion (the deterministic time scale) and the Kramers rate (i.e., the inverse of the average switch rate induced by the noise alone: the stochastic time scale); thus the term *stochastic resonance*. A comprehensive overview of stochastic resonance has been written by Gammaitoni et al. [14].

5.7 ON NON-SMOOTH SYSTEMS

New concepts that are important for the analysis of non-smooth systems in the state space will be introduced, and some of the possible bifurcations in non-smooth systems will be briefly surveyed in this section.

In non-smooth dynamic systems, the location of the discontinuities in the state space must be defined by additional functions $h_i(x_j) = 0$, $i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, N - 1$, where M is the number of discontinuities and N is the dimension of the state space. In the N-dimensional vehicle dynamic

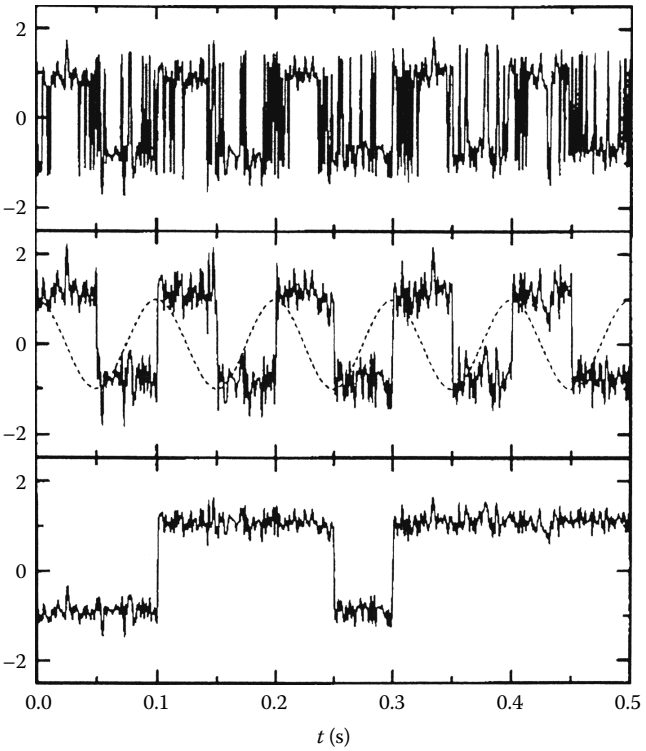


FIGURE 5.18 An example of input/output synchronization in a bistable system. The noise intensity is increasing from the bottom to the top with constant frequency. The dashed curve shows the input signal. (With permission from Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F., *Reviews of Modern Physics*, 70, 1, 223–287, 1998. Copyright 1998, by the American Physical Society.)

problems, the functions $h_i(x_j) = 0$ define *hypersurfaces*, each of which splits the state space into two subspaces. The hypersurfaces are the so-called *switching boundaries* and they will in general intersect each other. A simple example in the phase plane (x_1, x_2) is an elastic impact at $x_1 = 0$. When a trajectory in that phase plane hits the line $x_1 = 0$ in the point $(0, x_2)$, the velocity x_2 changes sign and the trajectory will continue from the point $(0, -x_2)$. The outgoing trajectory will return into the half plane it came from with a slope, which has the opposite sign of that of the incoming trajectory. Whenever a trajectory in the state space touches a switching boundary, an *event* occurs. A switching boundary may divide two domains in the state space with different DOF of the dynamic system. Systems with such a switching boundary are called *structure varying* systems. In Section 5.8, we shall outline a numerical procedure for the handling of the events. Further material is found in Leine and Nijmeijer [15] where numerical integration methods for the solution of non-smooth dynamic systems are presented.

Systems that have a continuous as well as a discrete nature are also called *hybrid systems*.

In nonlinear, non-smooth dynamic systems, the eigenvalues of the linearized systems may cross the imaginary axis in continuous dependence on the control parameter, but they may also jump across the imaginary axis from the negative to the positive real half plane, when an event occurs. In case the eigenvalues depend continuously on the control parameter, the bifurcations are created in the way we already described in Section 5.4, but the bifurcating solutions may depend differently on the control parameter in a neighborhood of the bifurcation point. Slivsgaard [16] found in her analysis of the dynamics of the prototype single-axle bogie for the Copenhagen S-train that the bifurcating stable, periodic solution will initially grow linearly with a finite slope from the bifurcation point. True [17] demonstrated in a simple example, that will be presented later, that the new behavior is generic for a group of non-smooth dynamic systems.

In their book, Leine and Nijmeijer [15] discuss bifurcations of nonlinear, dynamic systems, where the eigenvalues jump across the imaginary axis. They call the phenomenon a *discontinuous bifurcation*. When the eigenvalues of the linearized system depend continuously on a single control parameter in the neighborhood of the bifurcation point, then in the generic case only one real-valued or one pair of complex conjugate eigenvalues will cross the imaginary axis at a time. If however the eigenvalues jump in a discontinuous bifurcation, then more eigenvalues may jump at the same time, which may lead to different kinds of bifurcations. Some of the eigenvalues may even jump back again, so it is not at all easily recognized that they have jumped back and forth across the imaginary axis. Such cases may lead to new kinds of bifurcations. A similar situation occurs in maps. The analog here is a jump of the Floquet multiplier(s) across the unit circle. We refer to Leine and Nijmeijer [15] for more information.

Leine and Nijmeijer [15] also discuss the bifurcations that result from smooth approximations of the non-smooth systems. They demonstrate in simple analytic examples that *smoothing may change the character of the bifurcations*. This feature revives the discussion of "to which degree vehicle dynamics can be non-smooth." In an impact contact forces are produced, but should the impact be modeled as an ideally elastic one with an instantaneous infinitely large contact force—possibly with a coefficient of restitution—or are the contact forces modeled better by a dynamic law extending over a finite time interval? Both cases are nonanalytic in the sense that discontinuities are introduced in the system, but the discontinuities are not equally strong. In the first case, the velocity—the first derivative—has a finite jump; while in the second case, the velocity varies continuously while the acceleration—the second derivative—jumps. The second case is an example of a weaker "non-smoothness" than the first one. We shall not discuss modeling aspects here, but return to the topic in Section 5.8.

We continue this brief introduction with two simple examples of bifurcations of non-smooth systems. The first one is a non-smooth analog of the saddle-node bifurcation in Section 5.4. It is taken from the book by Leine and Nijmeijer [15]. Let us consider the non-smooth system:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lambda - |x|$$

There exist exactly two equilibrium solutions but only for $\lambda > 0$: $x = -\lambda$ and $x = \lambda$. This yields the bifurcation diagram shown in Figure 5.19, which should be compared with Figure 5.11a.

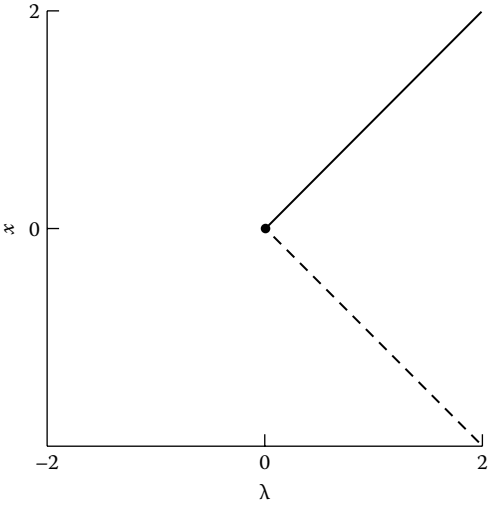


FIGURE 5.19 A discontinuous saddle-node bifurcation.

When the two figures are compared, we find that the stable and unstable solutions meet at a finite angle in the bifurcation point, when the system is non-smooth, and the solutions depend linearly on the control parameter λ. This is in contrast with the classical saddle-node bifurcation in Figure 5.11 (left). A stability analysis of the two branches shows that the stability properties are the same in the smooth and the non-smooth cases.

The other example is a non-smooth analog of the classical Hopf bifurcation shown in Figure 5.11 (right) in Section 5.4. True [17] considered the system:

$$\dot{x} = y$$

$$\dot{y} = -\left(k\sqrt{x^2 + y^2} - \lambda\right)y - x, \quad k \neq 0$$

We introduce polar coordinates by the transformation:

$$x = r\cos v, \quad y = r\sin v$$

and obtain

$$\dot{r} = (\lambda - kr)\, r\sin^2 v$$

$$\dot{v} = (\lambda - kr)\cos v \sin v - 1$$

This system has the solutions: $r = 0$ and $r = \lambda/k$.

A simple bifurcation analysis yields the bifurcation diagram in Figure 5.20. Here, we see that the bifurcating periodic solution $r = \lambda/k$ branches off linearly from the stationary solution $r = 0$, with a finite angle in contrast to the classical Hopf bifurcation. A stability analysis of the three branches shows—like in the preceding example—that the stability properties are the same in the smooth and the non-smooth cases. In both cases, the asymptotically stable stationary solution becomes unstable on the right-hand side of the bifurcation point. The periodic solution is asymptotically stable, when it exists for positive values of the parameter λ, and otherwise it is unstable.

Impacts and stick-slip caused by dry friction contact are the two sources of non-smoothness that occur most often in vehicle dynamics. In nonlinear dynamic systems, repeated impacts are considered to generate chaos in general. There is a lot of evidence hereof as well from experiments as from numerical simulations. When chaos is generated by repeated impacts, it is important to bear the nature of chaos in mind. An infinitesimal disturbance from a trajectory on a chaotic attractor will develop into a large departure from the values on the original attractor at a later point in time. This is the well-known effect of the sensitivity on initial conditions in chaos. The departed value will, however, lie on another trajectory on the same chaotic attractor. Since the uncertainties of measurements as well as the error bounds in numerical tests are finite, it means that in reality it is impossible

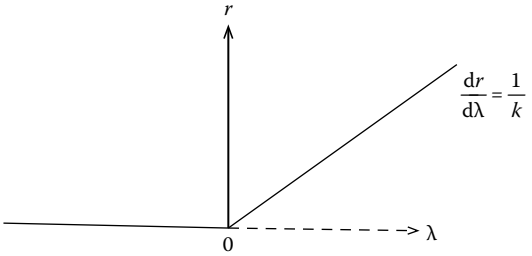


FIGURE 5.20 A discontinuous Hopf bifurcation.

to follow a trajectory on a chaotic attractor. The best we can hope for in any case is to shadow the chaotic attractor by careful measurements or suitable numerical algorithms. The modeling of the impact itself will therefore have a negligible effect on the general dynamics after the impact—as long as the dynamics stays on the same chaotic attractor. We shall return to this subject in the following section on numerics (Section 5.8).

This conclusion is also true for modeling of the stick in dry friction, but greater care must be exercised in that case. The stick has other important consequences for the dynamics. Stick will break some symmetries in the system. There is no proof of this, but the author has found this to be the case in all the vehicle dynamic systems his coworkers and he have analyzed. Thereby chaos is more likely to develop. Stick may also change stationary points in the state space into *set-valued stationary domains* and more generally equilibrium points into *set-valued equilibrium domains*. As an example of a stationary domain, let us consider a rolling railway wheelset that carries a car body in a suspension with dry friction damping. On straight track, the damper(s) may stick in positions that hold the wheelset fixed in a nonzero lateral displacement. The state is stable. The entire set of stable stationary lateral displacements and stationary yaw angels depends in general on the speed of the vehicle. In the case of the single rolling wheelset, the stable stationary domains will constitute a 3D hypersurface in the parameter-state space. The theory of bifurcations for smooth systems must be modified for dynamic systems with these properties. Such a wheelset may however, under different conditions, oscillate and reach equilibrium on one of several different cycles in the state space, each of which depends on the initial condition. These cycles in the parameter-state space constitute an equilibrium domain.

Leine and Nijmeijer [15] also discuss dry friction problems. Other examples of the special effects of dry friction damping on the dynamics of vehicles may be found in the following chapters.

Hoffmann et al. [18] investigated numerically the dynamics of a two-axle freight wagon with UIC standard suspension. The UIC suspension carries the car body in links with rolling dry friction in the contact surfaces between the elements of the links. Hoffmann et al. have calculated a bifurcation diagram, which has a hysteresis loop. The calculation was started with symmetric initial conditions, for example, the lateral displacement and the yaw angle of the wheelsets were zero. It is a stable stationary solution of the dynamic problem. After going around in the hysteresis loop, the stable stationary solution did not return to zero, but ended on a stable stationary nonzero solution. Their result is shown in Figure 5.21.

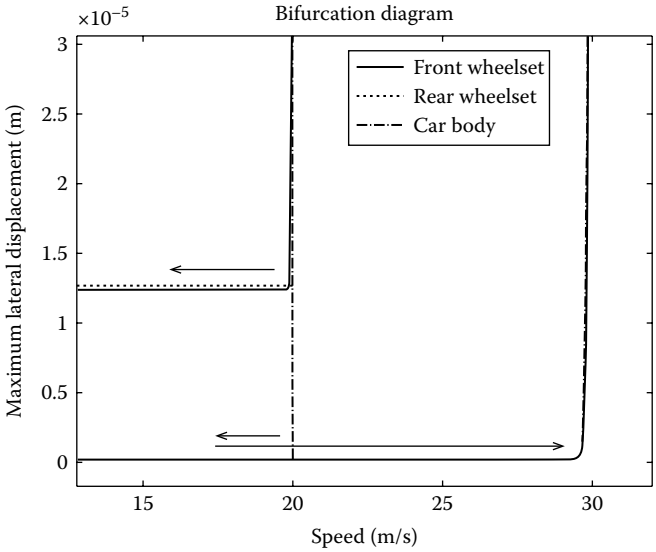


FIGURE 5.21 The "open hysteresis loop" in a model of a freight wagon with dry friction in the suspension verifying the existence of a stable stationary domain.

True and Asmund [13] investigated a numerical model of a single wheelset under a railway freight wagon. The model was simplified to illustrate only the effect of the coupling between lateral dry friction dampers with stick-slip and the creep forces. Therefore the kinematics of the wheel-rail geometry was linearized, and the wheels had no flanges. Figure 5.14 taken from [13] shows an interesting comparison between the bifurcation diagrams for the wheelset with no dampers and one with dry friction dampers. With no dampers the dynamic system can be reduced to a set of differential equations, and the mathematical theory in Section 5.4 applies. The bifurcation diagram is a classical one, with a trivial solution, which is asymptotically stable up to the Hopf bifurcation point, where a supercritical, asymptotically stable periodic solution branches off. The stationary trivial solution loses stability at the bifurcation point. For higher speeds, the asymptotically stable periodic solution goes through a symmetry breaking bifurcation and a series of—what looks like—period-doubling bifurcations, which lead to chaotic behavior. With the dry friction damper, the diagram is drastically changed. With the friction damper, the trivial solution is no longer asymptotically stable. We now find a set of stable symmetric periodic motions with amplitudes, which grow slowly with the speed, and which depend on the initial conditions. There exists now a whole range of stable periodic solutions with different amplitudes. Only one is shown in Figure 5.14. Its amplitude grows faster with the speed near the bifurcation point for the undamped bogie in a way that is known from nonlinear systems with noise. The Hopf bifurcation has disappeared.

Leine and Nijmeijer [15] write two chapters with "Discussion and Conclusions." We shall try to condense the conclusions relating to bifurcations that are most important for vehicle dynamics. For a deeper discussion, the reader is referred to [15] or to the review paper [19].

There exist types of bifurcations in non-smooth dynamic systems that do not have a smooth counterpart, and the mathematical theorems on bifurcations are almost all based on assumptions of some smoothness. Therefore, *conclusions about the existence of bifurcations must be taken from inspection of bifurcation diagrams.* Unfortunately, we are in general unable to find the unstable solution branches numerically; therefore, the computed bifurcation diagrams will be incomplete. Leine and Nijmeijer [15] analyze examples of bifurcations from a stationary solution in non-smooth systems. They find by analytic calculations that an eigenvalue of the Jacobian has crossed the imaginary axis. There exists, however, no mathematical proof that a crossing is a necessary condition for a bifurcation in a general non-smooth system. In our own numerical investigations of vehicle dynamic systems, we found that the crossings of eigenvalues that are known from smooth systems also take place in our non-smooth systems. We found it to be true for bifurcations from stationary solutions as well as from limit cycles. See the review paper [19] for more information.

The conclusion is that *it is necessary to calculate bifurcation diagrams in order to determine the bifurcation points in vehicle dynamic problems.* This leads to the next section in which a simple numerical method for the determination of the *stable* branches of a general nonlinear dynamic problem will be described.

5.8 ON THE NUMERICAL ANALYSIS OF VEHICLE DYNAMIC SYSTEMS

5.8.1 Numerical Solution of Smooth Problems

In this section, we shall briefly survey the numerical solution of vehicle dynamic problems. Special attention will be paid to problems with coexisting attractors and non-smooth problems. We shall demonstrate how such problems must be treated in order to obtain reliable dynamic results.

When coexisting attractors exist in a nonlinear dynamic problem, the equilibrium solutions of a specific initial value problem depend on the initial values. This is in contrast to linear dynamic problems, where coexisting attractors cannot exist. On the other hand, the solution of a given initial value problem is unique provided certain continuity conditions are satisfied. In other words, the trajectory in the state space is uniquely defined. If, however, two trajectories that converge toward two different attractors are sufficiently close in some part of the state space, then numerical time step integration

along one of the attractors may jump to and follow the other attractor. This leads to an erroneous numerical solution of the initial value problem. The jump may be the result of a bad choice of error bounds or time steps, but it may also be the result of the use of the wrong numerical routine. We shall demonstrate later the erroneous result of the application of an Euler explicit routine (which should NEVER be applied to nonlinear equations [!]) to a simple nonlinear initial value problem.

In Section 5.2, we solved

$$2.56\,\mathrm{d}^2\frac{x}{\mathrm{d}t^2}+0.32\,\frac{\mathrm{d}x}{\mathrm{d}t}+x+0.05x^3=2.5\cos t \tag{5.7}$$

numerically with two pairs of initial conditions. Two different attractors were found, and they are shown in Figure 5.1.

Let us now consider Equation 5.7 with the initial condition

$$x(0)=3.75,\quad \frac{\mathrm{d}x}{\mathrm{d}t(0)}=0.75 \tag{5.8}$$

We first solve the initial value problem (5.7 + 5.8) with a ode45 routine from MAPLE 9.5. The transient approaches the correct attractor with small amplitude that belongs to the initial conditions (5.8). The result is shown in Figure 5.22.

We next solve the same initial value problem (7 + 8) with an explicit Euler routine—also from MAPLE 9.5. The result is shown in Figure 5.23, and it is clearly seen that *the transient in this case approaches the other—and incorrect—attractor.*

The explicit Euler routine is very popular, because it works fast. As long as no other solution or measured results from a test exist for comparison, the Euler solution will be accepted. It is outright dangerous. The explicit Euler routine should never be used for the numerical solution of nonlinear dynamic problems.

This simple example demonstrates that great care must be taken, when a numerical solver is chosen for a given nonlinear dynamic problem.

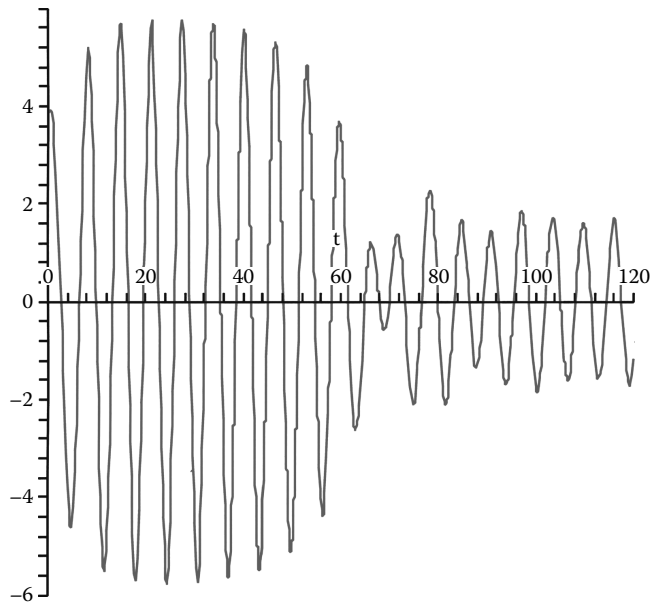


FIGURE 5.22 The numerical solution to the initial value problem (7) + (8) using a Runge–Kutta 45 routine. The transient approaches the correct attractor.

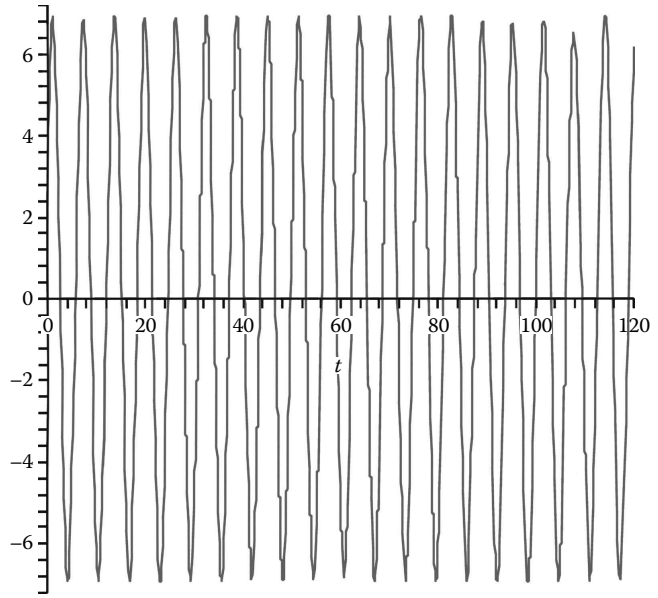


FIGURE 5.23 The numerical solution to the initial value problem (7) + (8) using an explicit Euler routine. The transient approaches *the wrong attractor.*

Garg and Dukkipati [20] discuss and demonstrate the use of explicit and implicit routines for the numerical solution of linear as well as nonlinear dynamic problems. They test a selection of routines on both linear and nonlinear dynamic problems. They compare CPU-time, stability, and artificial damping in their examples. In an *explicit* routine, the next step $\boldsymbol{x}_{n+1}$ depends explicitly on known quantities only. In Garg and Dukkipati's nonlinear example [20] with the rolling wheelset, the explicit solvers either are unstable or introduce artificial damping. The explicit solvers are inefficient for the analysis of the so-called stiff problems. Generally, a problem is said to be *stiff*, if the solution contains both slow and very fast processes. The fast processes decay very fast with time t, so a very short step length is needed to keep the explicit solvers from "blowing up" in the numerical solution. Furthermore, the explicit solvers often become unstable when they are applied to stiff problems. In order to maintain stability, the step size again needs to be very small. In *implicit* routines, the dependence is implicit, so the next step has to be calculated by the solution of an equation. Since vehicle dynamic problems are nonlinear, the said equation will also be nonlinear. It means that a nonlinear equation has to be solved numerically for each time step for example by a *predictor corrector method*. The advantages of the implicit methods in vehicle dynamics are that they can be applied safely to solve stiff problems with automatic control of the varying step lengths. Vehicle dynamic problems as well as the numerical simulations of real-world tests are usually stiff problems. Therefore, the application of implicit routines is recommended for the numerical solution of these initial value problems.

Rill [21] has applied an implicit Euler routine to vehicle dynamic problems with success. True and his coworkers have most often used the LSODA routine for the calculation of bifurcation diagrams—also successfully. LSODA changes solution strategy automatically depending on whether the problem is stiff or not. The Runge–Kutta solver ode45 has been recommended for years as the most effective routine for the purpose of shadowing chaotic attractors, but it is an explicit routine, and it therefore cannot be recommended for the numerical analysis of vehicle dynamic problems. There exist, however, also implicit Runge–Kutta solvers for stiff problems. All these Runge–Kutta solvers are a part of MATLAB. Most recently, the routine SDIRK [22], based on the theory from [23], has been applied by Hoffmann [24] to the numerical analysis

of two-axle freight wagons—a stiff problem with rolling and sliding dry friction and bumper stops. SDIRK is a recent implicit routine, which can be downloaded from the web. For a detailed description of time-stepping explicit and implicit numerical routines and recommendations for their applications, the interested reader is referred to the literature [25,26].

The time-stepping numerical methods are by far the most popular numerical methods for the solution of vehicle dynamic problems, because they lend themselves naturally to the solution of initial value problems. It should be mentioned, however, that dynamical problems can be formulated as variational problems, where a supremum of a functional is sought for example using Hamilton's principle. This formulation leads to an optimization problem in mathematics. Equilibrium solutions have been found by numerical solution of the optimization problem. It is claimed (oral information) that the stable attractors in a bifurcation diagram for a vehicle can be computed faster in that way.

The engineer without a good background in numerical analysis faces a problem when he must choose a differential equation solver (DE solver). Although a numerical solver may deliver a solution when it is applied to a given dynamic problem, it cannot be taken for granted that the solution is the correct one. It is strongly recommended that the engineer applies two different and acknowledged routines to the problem—at least initially. If the result using one of the solvers differs considerably from the results, which the other solver yields, he must conclude that at least one of the solvers is bad in his case. On the other hand, he cannot be sure that the results are correct even if the two results agree with each other.

5.8.2 Numerical Solution across a Discontinuity

When a vehicle dynamic problem is not smooth, the dynamic problem must be supplemented with functions $h_i(x_j) = 0$, which define the location of the discontinuities in the state space. The functions are *hypersurfaces* in the state space (see Section 5.7). By the word "discontinuities" is understood not only discontinuities in the functions themselves (zero-order discontinuities) but also discontinuities in the derivatives up to some order (higher-order discontinuities). The hypersurfaces are the so-called *switching boundaries*. As an example, we consider now a dynamic problem with one switching boundary $h(x_j) = 0$, $j = 1, 2, \ldots, N - 1$, where N is the dimension of the state space. The analysis is easily generalized to problems with more switching boundaries as long as we can deal with them one at a time. In the two domains, divided by the switching boundary, the problem is smooth, and thereby we solve the initial value problems in the usual manner. All we, in addition, need to consider is the region close to the point P, where the solution hits the switching boundary. The point P is called the *transition point*.

When we integrate our dynamic problem as an initial value problem in a domain where the dynamic system is smooth, we must stop the integration as soon as the trajectory crosses a switching boundary. The switching boundary is only defined in terms of the *dependent variables* $x_j(t)$. We must therefore continuously check in our solution procedure at which time t_0 the trajectory has crossed the switching boundary. It is then necessary to reformulate the solution procedure to determine time T iteratively at which the trajectory hits or touches the switching boundary. T can only be found numerically within a certain error limit, but it should be determined as accurately as possible. If an implicit solver is used, the interpolant may be used to determine T using Newton–Raphson's method. An initial guess is derived directly from the linear interpolant between the two solution points surrounding the event. We have now found the transition point P. It is defined by the values $x_j(T)$, $j = 1, 2, \ldots, N$.

In the mathematical formulation of the non-smooth dynamic problem, it must be defined what will happen at the switching boundary. The collision with the switching boundary and what happens there is called an *event*. As an example of an event, we mention a jump Δv in the first derivative of one of the state variables—say x_K—at time T.

We then formulate a new initial value problem at time T using the end-state variables found at the transition point P, adjusted by the law of the event, the new initial values for the integration. In our

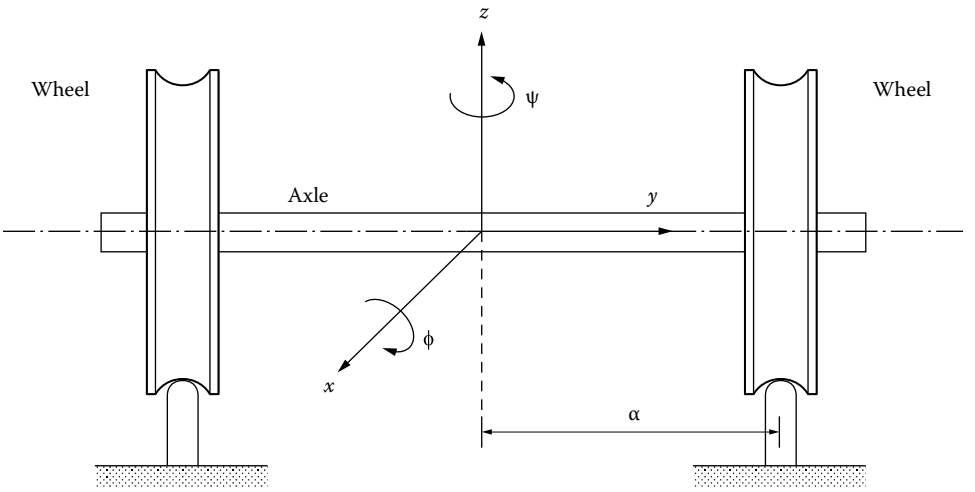


FIGURE 5.24 The wheelset model.

example, we take all our end-state variables at time T and only add the given jump Δv to the derivative of x_K at time T. The continuation of our dynamic problem is then given by the new initial values at time T combined with the smooth dynamic formulation of our problem in *the relevant domain*. In our example, the relevant domain will be *on the other side of the switching boundary*—the second domain. Then the integration continues in that domain.

As an example, we investigate the dynamics of a wheelset used on a railway wagon and presented in the book by Garg and Dukkipati [20]. The system is shown in Figure 5.24, where the wheelset is traveling along a track.

The nonlinear equations of motion for the lateral displacement and the yaw angle and the values of the different physical parameters can be found in [20].

The variation of the difference in the rolling radii of the wheelset $\frac{r_L - r_R}{2}$ with lateral displacement is shown in Figure 5.25.

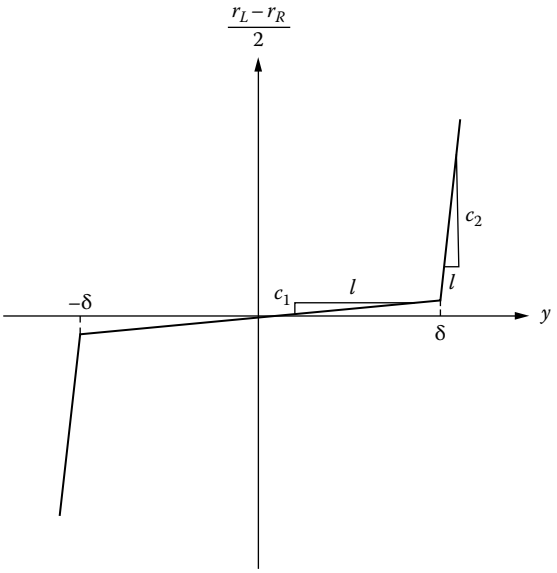


FIGURE 5.25 The variation of different variables with lateral displacements.

For different entries, we have the following values of the constants c_1 and c_2; the distance $\delta = 0.35$ in the experiments:

var	c_1	c_2
Δ_1	0	18
Δ_2	0	18
Δ_L	0.15	22
$\frac{r_L - r_R}{2}$	0.05	10

Solving the problem for an initial lateral displacement $y = 0.36$ with a velocity of 160 km/h leads to the solution In Figure 5.26, where the damping is strong and only one passage of the discontinuity occurs. Here, the solution with constant step size called ONERK4 and the solution that tracks down the discontinuity called GERK are very close.

When the velocity is increased to near the Hopf bifurcation point of this *purely theoretical problem*, where more passages of discontinuities occur, we find very different solutions. Figure 5.27 demonstrates that the constant step size solution is less accurate than the one where the discontinuities are tracked. This leads to different interpretations of the dynamic properties. This may be critical since this case is near the point where the stationary solution becomes unstable.

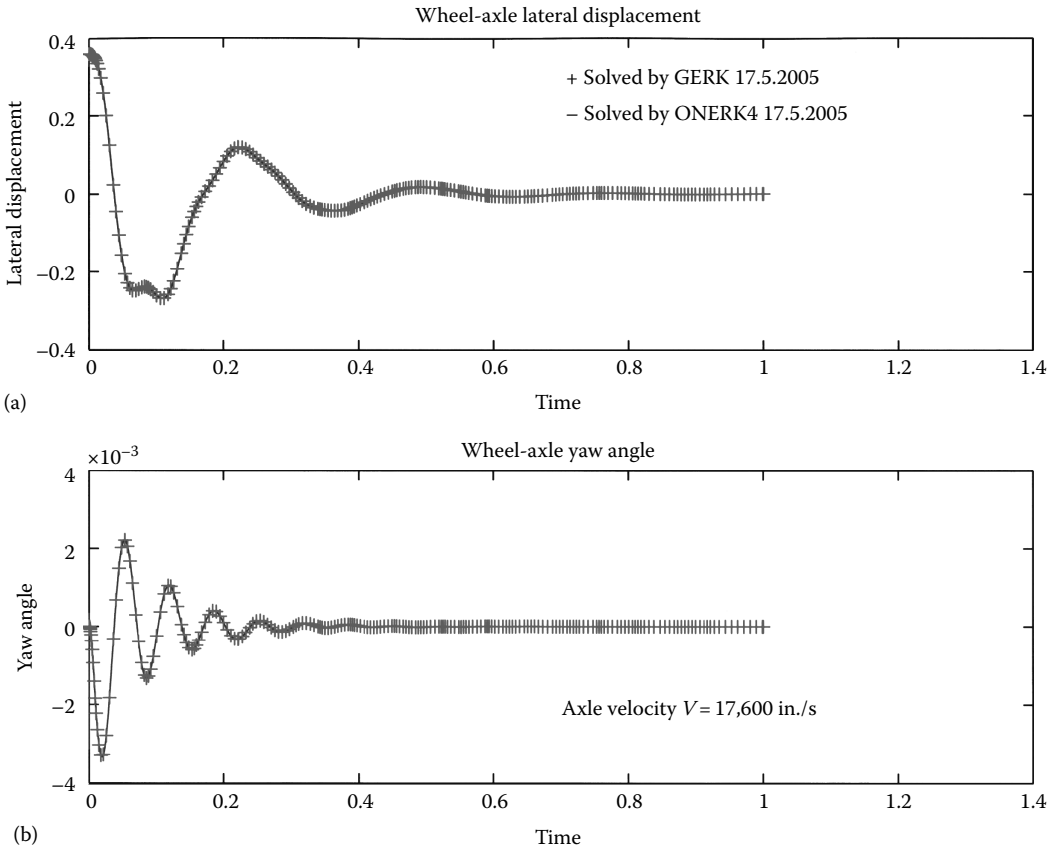


FIGURE 5.26 Wheelset dynamics solution for $V = 160$ km/h.

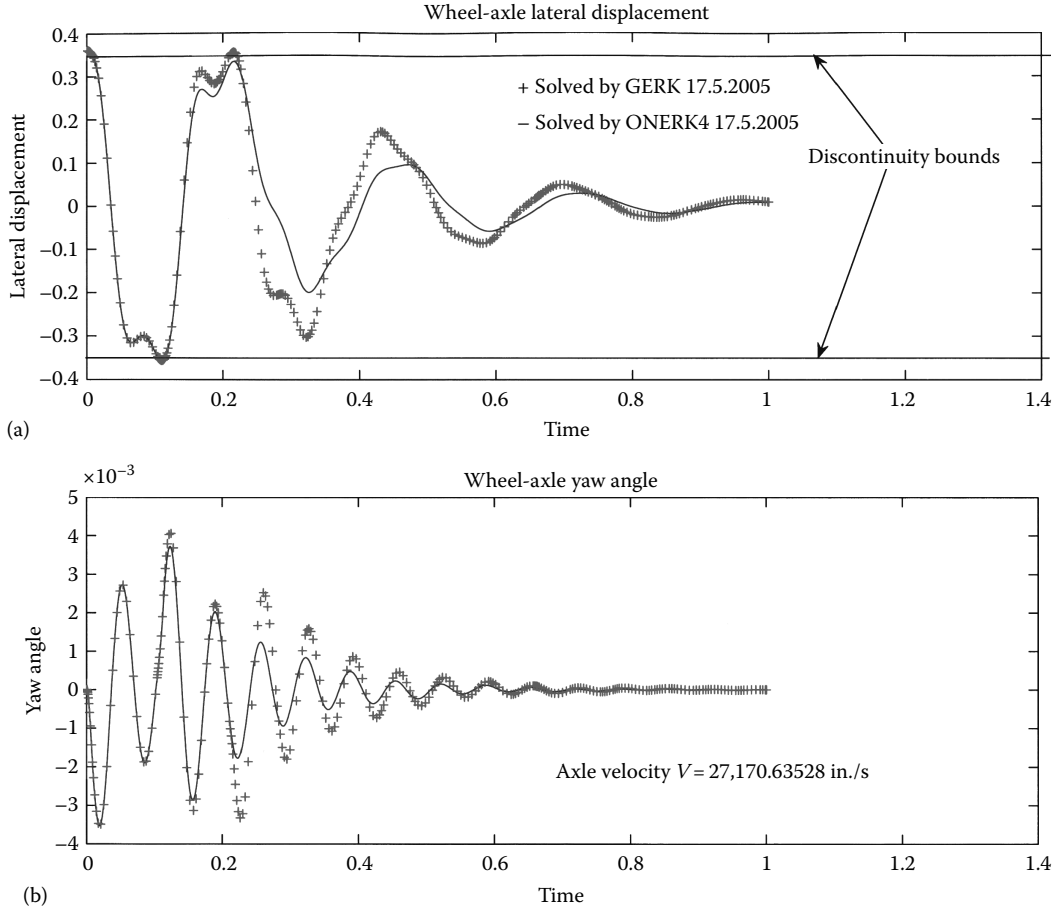


FIGURE 5.27 Wheelset dynamics solution for $V = 2484$ km/h.

Let us summarize our experience gained from different numerical treatments of discontinuities in vehicle dynamic problems. The simplest way to handle the discontinuities is to *ignore the discontinuities and integrate across them as if they do not exist*. One hopes that the "smoothing property" of the numerical routine will automatically take care of the problem with the discontinuity. If the discontinuity is in the second or higher derivatives of the functions in the dynamic problem, this strategy may work well. The strategy might work also in other cases. The general attitude is that as long as the numerical integration delivers a solution then the solution is "the correct one." This is rarely true, as we just demonstrated in the last example. A solution found in this way may even be qualitatively wrong. A chaotic dynamic behavior may be totally missed. This was illustrated by Xia and True [27]. Last but not least, the numerical solver may use a lot of CPU-time with the integration across the discontinuity, and we do not know the resulting numerical approximation of the smoothing function.

The problem with the resulting smoothing function leads to the second method of handling the discontinuities: *The smoothing functions are prescribed in each and every case*. Then—at least—we know how the numerical routine handles the discontinuity, and the CPU-time will be reduced. The smoothing function must model the physical reality in the best possible way. Many people claim that this method is the optimal one, because it models the physical reality closely. The problem with the method is, however, that "a good mathematical model" is not always known. In the work by True and Asmund [13] and True and Trzepacz [28], a smooth model of stick-slip dry friction in plane contact between steel or cast iron surfaces was applied. The results of the numerical dynamic

analysis contain a lot of chaotic dynamics as could be expected from the nature of the problem. In True and Trzepacz [28], the impacts were not modeled with smoothing functions. A comparison was, however, made between an ideally kinematic model and a dynamic model of an elastic impact. The results differed—as could be expected—but both results lay on the same chaotic attractor. Only the last model is able to yield an estimate of the contact forces. True and Trzepacz [28] used a constant step length in his integration. Only in the case of the dynamic model did he interpolate to determine the location of the impact with great accuracy. In the kinematic case, the time of reversal was simply found by a calculation of the position of the moving part. When the moving part had "entered" the part in rest, the velocity of the moving part was reversed. This approximation delivered satisfactory dynamics as long as the step length was sufficiently small. In the dynamic model, the CPU-time used for integration across the impact exploded. It is interesting to note, however, that the dynamics of the system is the same chaotic one in the two cases. The situation was discussed in Section 5.7 with relation to the modeling of the impact. The robust dynamics is closely connected with the chaotic dynamics in this case. The robustness is however not guaranteed. The dynamics of the two cases may differ.

The third—and correct—way of handling the discontinuities numerically is to *track down the transition point as accurately as possible.* The transition point is the point in the state space where the trajectory hits the switching boundary. *Then stop the time integration of the dynamic problem at that point, apply the law of transition, and continue the time integration of the dynamic problem on the relevant side of the switching boundary.* This is the safest and most accurate method to deal with nonsmooth problems. The extra CPU-time spent in the determination of the transition point is balanced by the time not spent for the integration of the dynamic problem with very small time steps along steep gradients across the discontinuity. It is the logical numerical application of the results presented in Section 5.7. The method was used with success by Hoffmann [24] to investigate the difficult problem of the dynamics of two-axle freight wagons. His models contain both rolling and sliding friction in the suspension as well as possible impacts between the axle boxes and their guidance.

The lack of proper consideration of the discontinuities in the numerical analysis leads in general to an even qualitatively erroneous result of the calculations. It is important to identify coexisting attractors in nonlinear dynamic systems. The necessity is obvious, when the attractors are characterized by different amplitudes. It may seem less obvious if the attractors have almost the same amplitudes and are only qualitatively different. If however—like in [13,27,28]—one of the attractors is a chaotic attractor, then great caution must be exercised. We refer to the discussion of chaotic transients at the end of Section 5.3. The existence of a coexisting chaotic attractor will often give rise to violently oscillating transient motions with dangerously high amplitudes. This happens for transients approaching either attractor as well the chaotic attractor as a coexisting non-chaotic attractor. It is therefore very important to identify all the chaotic attractors in a nonlinear dynamic problem. The transients may lead to accidents in the physical world.

5.8.3 A Way to Compute Stable Branches of Vehicle Dynamic Problems Numerically

In order to find the stable branches—the attractors—of a vehicle dynamic problem, use *path following in parameter-state space*. Given a stable equilibrium solution to our dynamic problem $\boldsymbol{X}(t)$ for a certain value of the control parameter, in our case this is the speed, V. We then know the solution at a time t_0 and store the state vector $\boldsymbol{X}(t_0)$. It is called the end value. The path following is performed stepwise. In each step, we increase or decrease V with a small increment and calculate the equilibrium solution of the resulting dynamic problem using the end value of the preceding step as the initial value for the new step. We take as many steps as needed. The process will end when the transient in the integration approaches another stable equilibrium solution for a certain value V_0 of the parameter V. V_0 is close to a bifurcation point that has just been passed, and which can now be determined by interpolation of the parameter V. In this way, a solution of the dynamic problem can be traced or *followed* and bifurcation points can be found in the state parameter space. At the

bifurcation points, we turn our attention to the new stable branches and follow them in dependence on V. It is highly recommendable to calculate the eigenvalues of the dynamic problem, linearized around the stable equilibrium solution and find their change with the parameter across the bifurcation point. The calculation is a valuable verification of the properties of the found bifurcation. In a bifurcation of a periodic solution from a stable stationary solution, two complex conjugate eigenvalues should cross the imaginary axis at the bifurcation point from the negative to the positive real half plane.

The success of the method is based on the assumption that the stable branches can be found as bifurcations from at least one stable solution (branch) of the problem that is known a priori. It must be emphasized that it is not guaranteed that *all* the stable branches can be found in this way in all dynamic problems. In vehicle dynamics, the aim is to design a vehicle that runs smoothly along a given track. We shall therefore make the basic assumption that the vehicle dynamic model to be investigated has a known stationary solution in a certain speed range in an appropriate moving coordinate system. In vehicle dynamics, the speed, V, is an appropriate control parameter, and in most cases "the appropriate coordinate system" is one that moves with the constant speed, $V > 0$ along the track of the vehicle. A dynamic model of a "good" vehicle will then have a trivial stable solution for sufficiently small values of the speed $V > 0$, that is, all $x_j(t) \equiv 0$. It can be used as the initial solution.

The path following procedure can be accelerated and automated by application of *the ramping method*. In the dynamic system, the control parameter V is substituted with a slowly varying monotonically growing or decaying function of time $V(t)$. In the ramping method, we start the integration of the dynamic problem as usual with the known end value and integrate up to an appropriate end time T, so that $[V(0), V(T)]$ covers the designed speed range of the vehicle. If bifurcations appear, repeat the procedure as long as necessary to determine all the attractors. The bifurcation diagram will in practice be quite well approximated by the ramping solution, but the method overshoots the bifurcation points, and one or more of closely situated bifurcation points may be missed. The bifurcation points must afterward be recalculated with greater accuracy.

The smallest parameter value V_C in a vehicle dynamic problem for which there exist more than one stable equilibrium solutions (coexisting attractors) to the dynamic problem is the *critical speed*. It is important to realize that the determination of the critical speed in general leads to a *mathematical existence problem*. In railway dynamic problems, there often exists an asymptotically stable periodic solution coexisting with the asymptotically stable stationary solution for lower speeds than the one at which the stationary solution loses stability. The periodic solution is then the result of a subcritical bifurcation of an unstable branch that gains stability in a saddle-node bifurcation. Since it is still impossible to path follow an unstable periodic solution in a differential-algebraic dynamic system, a different strategy must be used to determine the critical speed, [32,33,34].

The calculation of V_C may be performed in three steps:

1. Substitute the known steady-state solution into the vehicle dynamic problem, increase the speed V, and solve the initial value problem for a very small disturbance of the initial values, for example, for a lateral displacement of one axle of 0.1 mm and all other initial values equal to zero. Then the steady-state solution is found for a higher value of the speed. Repeat the procedure for higher and higher values of the speed until the solution of the initial value problem no longer is the known steady-state solution and stop the calculation when the transients have vanished. Store the end values of all the state variables, $\boldsymbol{X}_{e1}$. We have now followed the known steady-state solution in parameter-state space until it lost its stability. Since the known steady-state solution is now unstable, the numerical solution has converged to the nearest asymptotically stable periodic solution in state space—a new attractor.
2. We must now follow the periodic solution for decreasing speed until it ceases to exist. Reduce the speed a little and solve the vehicle dynamic problem with the end values $\boldsymbol{X}_{e1}$ as the new initial conditions. Stop the integration routine when the transients vanish and store the new end values $\boldsymbol{X}_{e2}$. We have now calculated another point on the periodic attractor.

Repeat the procedure for lower and lower values of the speed until the periodic solution ceases to exist. We have then just passed the saddle-node bifurcation point of the periodic attractor and are therefore close to the theoretical critical speed.

3. We can now determine the saddle-node bifurcation point, that is, the theoretical critical speed V_C with a desired accuracy by interpolation of the parameter V or by the Newton–Raphson's method.

Other authors have suggested other methods for the calculation of the critical speed. These other methods are discussed in an important work by True [29].

Continuation algorithms based only on Jacobians are not recommended, because we already know a solution of the full nonlinear dynamic problem. When an implicit numerical solver is applied, then the Jacobian will be calculated and updated automatically in every time step.

There exist several good commercial vehicle dynamic simulation routines. When a potential user must select a program, he must test his requirements of the program against the performance of the program and against the requirements listed in this section. The programs must contain at least two recognized implicit solvers for the solution of nonlinear dynamic problems, and the user must be able to control the numerical parameters such as step lengths—constant or varying size—and error bounds. Explicit Euler routines are of course banned from use.

If not treated properly, the interpretation of the dynamic properties from the numerical results may be very different from the properties of the real system.

We recommend when the numerical tools are selected, that a similar care is spent on the selection of the code as the care taken of the modeling. It is of the at most importance that properties that influence the dynamic behavior such as the discontinuities are properly accounted for.

5.9 A GUIDE TO VEHICLE SYSTEM DYNAMICS

Due to the nonlinear coupling with the supporting ground, it is basically impossible to split the dynamics of the total vehicle-ground system into effects caused by the properties of the vehicle and those of the support. The contact forces between the ground-rail and wheels are important unknowns in the simulations not only because they act as an input to the dynamic vehicle model but also because they are important for the design of the vehicle and the deterioration of the structure of the support. We are, however, lucky in the sense that the influence of the dynamic feedback from the vehicle on the ground is on a much larger time scale than the instantaneous action the vehicle is exposed to from the ground. The feedback can therefore be neglected in the modeling of the short-term dynamics of the vehicle.

It is, however, possible to make deductions and achieve a better understanding of the dynamics of the total system if the dynamics of some of the elements in the system is known. The dynamic features of the vehicle model should therefore be investigated first. The investigation of the full nonlinear vehicle dynamic model starts with a numerical calculation of the bifurcation diagrams for the vehicle running on an ideal support along a straight line (Section 5.8.3). The calculations must be performed for all relevant tire brands or contact geometries and for a relevant selection of adhesion coefficients. The speed is chosen as the control parameter, while all other parameters are kept constant in each run. The relevant attractors in the parameter-state space must be found and the critical speed determined. Next, the investigation continues with a set of numerical calculations of the bifurcation diagrams for quasi-stationary curving in arcs with selected radii. Again the speed is chosen as the control parameter. A lower critical speed will often be found in curves with a large radius than the one found by driving along a straight line.

When the attractors are known, we are ready to make numerical simulations of the vehicle dynamics on uneven roads, transition curves, turnouts, and in other situations where the transient response is important. These simulations also serve as a basis for comparisons with measured results from road tests and as a verification of the dynamic model. This may give rise to a great variety of nonlinear motions. In his master thesis, Christiansen [30] investigated the dynamics of the

Cooperrider [31] bogie on a sinuous track. Depending on the phase difference between the amplitude of the sinus functions of the right and left rail, Christiansen found many nonlinear phenomena known from the literature—but only when the speed was above the critical speed. The phenomena were obviously results of mode interactions with the hunting motion of the vehicle. Since most railway freight vehicles today often run at supercritical speeds, special care must be taken when their dynamics is modeled and investigated.

In every step of the investigation, it is important to pay attention to the numerics. After having spent so much effort on the construction of the theoretical model, many investigators seem to believe that the job is done. All that is needed at that point is not just any numerical solver, which can be found on the web, or which is incorporated in a commercial simulation program—often without any information about the type of solver and its parameters. We recommend, when the numerical tools are selected, that the choice is made with a care similar to the care invested in the modeling. It is of the outmost importance that properties such as the discontinuities that influence the dynamic behavior are properly accounted for (see Section 5.8.2).

It is, of course, permitted to make approximations of solutions to nonlinear systems for instance by linearization of nonlinear characteristics or terms in the dynamic system. If the total dynamic system will be linearized, it is important that the operator knows which stable solution, that is, *stable branch* he will linearize around. He must also investigate whether a linearization is permitted (see Section 5.3.3). The Jacobian $\boldsymbol{J} = [\partial f_i/\partial x_j]$, $i, j = 1, 2, \ldots, N$ must exist, and the rest term $\boldsymbol{G}(\boldsymbol{x}; \boldsymbol{\lambda})$ must satisfy $\|\boldsymbol{G}(\boldsymbol{x}; \boldsymbol{\lambda})\|/\|\boldsymbol{x}\| \to 0$ for $\|\boldsymbol{x}\| \to 0$ if the trivial solution is considered. It is therefore important to know the full nonlinear operator as well as its bifurcation diagram first. The linearized operator can yield valuable information about certain properties of the vehicle model. The eigenvalues will for instance inform about characteristic frequencies and their attenuation and potential resonances in the vehicle dynamic system.

When the analyses of the nonlinear dynamics of vehicle systems are performed with accurate theoretical multi-body system models that consist of elements with well-defined and verified dynamics, with due regard to the nonlinear nature of the systems, and the calculations are performed on modern computer systems by well-educated personnel using software that is carefully selected with regard to the nature of the specific dynamic problem, then the results will be so accurate that simulations in future can be used as a documentation for certain vehicle performance requirements and even as a part of the safety specifications. This will give the manufacturers, the national boards of safety and large user organizations—like the railway companies—a strong and reliable tool for the design and testing of vehicle systems under realistic or even unrealistic operating conditions. The models may also be valuable tools for investigations of accidents. The advantages of such theoretical analyses are lower costs compared with alternative methods, flexibility with respect to design parameters and operating conditions, and—in many cases—a faster production of results.

ACKNOWLEDGMENTS

The author thanks Railnet Denmark and Banverket Sweden for their financial support. Per Grove Thomsen and Mark Hoffmann are acknowledged for reading the manuscript and for their valuable criticism.

REFERENCES

1. Thompson, J. M. T. and Stewart, H. B. *Nonlinear Dynamics and Chaos*, John Wiley & Sons, New York, 1986.
2. Moon, F. C. *Chaotic and Fractal Dynamics*, John Wiley & Sons, New York, 1992.
3. Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer Verlag, Berlin, Germany, 1983.
4. Jordan, D. W. and Smith, P. *Nonlinear Ordinary Differential Equations*, Clarendon Press, Oxford, U.K., 1988.

5. Marek, M. and Schreiber, I. *Chaotic Behaviour of Deterministic Dissipative Systems*, Cambridge University Press, Cambridge, U.K., 1991.
6. Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics*, John Wiley & Sons, New York, 1995.
7. Ott, E. *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge, U.K., 1993.
8. Wiggins, S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer Verlag, Berlin, Germany, 1990.
9. Wiggins, S. *Global Bifurcations and Chaos*, Springer Verlag, Berlin, Germany, 1988.
10. Pascal, J.-P. Oscillations and chaotic behaviour of unstable railway wagons over large distances, *Chaos, Solitons and Fractals*, 5, 1995, 9, 1725–1753.
11. Maicz, D., Mittermayr, P., Presle, G., and Stephanides, J. Muster im chaotischen dynamischen System Eisenbahn—Bildung von Strukturen, ZEVrail Glasers Annalen, 132, Tagungsband SFT Graz 2008, 2008, pp. 190–201.
12. Øksendal, B. *Stochastic Differential Equations—An Introduction with Applications*, Springer Verlag, Berlin, Germany, 1989.
13. True, H. and Asmund R. The dynamics of a railway freight wagon wheelset with dry friction damping, *Vehicle System Dynamics*, 38, 2002, 2, 149–163.
14. Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F. Stochastic resonance, *Reviews of Modern Physics*, 70, 1998, 1, 223–287.
15. Leine, R. I. and Nijmeijer, H. *Dynamics and Bifurcations of Non-Smooth Systems*, Springer Verlag, Berlin, Germany, 2004.
16. Slivsgaard, E. C. On the interaction between wheels and rails in railway dynamics, Ph.D. Thesis, IMM, The Danish Technical University, Kgs, Lyngby, Denmark, 1995.
17. True, H. *On a new Phenomenon in Bifurcations of Periodic Orbits, Dynamics, Bifurcation and Symmetry, New Trends and New Tools*, September 3–9, 1993. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1994, pp. 327–331.
18. Hoffmann, M., Petersen, D. E., and True, H. On the dynamics of a railway freight wagon with UIC Standard Suspension, *Proceedings of the 7th Conference on Dynamical Systems—Theory and Applications*, Technical University of Łódz, Łódz, Poland, Vol. 1, December 8–11, 2003, pp. 135–160.
19. di Bernardo, M., Budd, C., Champneys, A. R., Kowalczyk, P., Nordmark, A., Olivar, G., and Piiroinen, P. T. Bifurcations in non-smooth dynamical systems, *SIAM Review*, 50, 2008, 629–701.
20. Garg, V. K. and Dukkipati, R. V., *Dynamics of Railway Vehicle Systems*, Academic Press, Toronto–Orlando–San Diego–New York–London–Montreal–Sydney–Tokyo, 1984.
21. Rill, G. A modified implicit EULER algorithm for solving vehicle dynamic equations, *Multibody System Dynamics*, 15, 2006, 2, 1–24.
22. Nørsett, W. A. and Thomsen, P. G. Imbedded SDIRK-methods of basic order three, *BIT*, 1984, 24, 634–646.
23. Alexander, R. Design and implementation of DIRK integrators for stiff systems, *Applied Numerical Mathematics*, 2003, 46, 1–17.
24. Hoffmann, M. On the dynamics of European two-axle railway freight wagons, *Nonlinear Dynamics*, 52, 2008, 301–311.
25. Lambert, J. D. *Computational Methods in Ordinary Differential Equations*, John Wiley & Sons, New York, 1972.
26. Öderlind, G. The automatic control of numerical integration, *CWI Quarterly*, 1998, 11, 1, 55–74.
27. Xia, F. and True, H. On the dynamics of the three-piece-freight truck, RTD-Vol. 25, *IEEE/ASME Joint Rail Conference*, Chicago, IL, American Society of Mechanical Engineers, United Engineering Center, New York, April 22–24, 2003, pp. 149–159.
28. True, H. and Trzepacz, L. The dynamics of a railway freight wagon wheelset with dry friction damping in the suspension, *Proceedings of the 18th IAVSD Symposium on Vehicle System Dynamics*, 2003, The Dynamics of Vehicles on Roads and Tracks, Taylor & Francis Group, London, U.K., 2004, pp. 587–596.
29. True, H. Multiple attractors and critical parameters and how to find them numerically: The right, the wrong and the gambling way, *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, 51, 2013, 3, 443–459.
30. Christiansen, L. E. The dynamics of a railway vehicle on a disturbed track, M.Sc. Thesis, Department of Physics, The Danish Technical University, 2001. http://orbit.dtu.dk/getResource?recordId=222571&objectId=1&versionId=1
31. Cooperrider, N. K. The hunting behavior of conventional railway trucks, *ASME Journal of Engineering and Industry*, 94, 1972, 752–762.

32. True, H. Does a critical speed for railroad vehicles exist? RTD-Vol. 7, *Proceedings of the 1994 ASME/IEEE Joint Railroad Conference*, Chicago, IL, American Society of Mechanical Engineers, United Engineering Center, New York, March 22–24, 1994, pp. 125–131.
33. True, H. On the theory of nonlinear dynamics and its applications in vehicle systems dynamics, *Vehicle System Dynamics*, 31, 1999, 5–6, 393–421.
34. True, H. Nichtlineare Schienenfahrzeugdynamik, neue Grundlagen, Methoden und Ergebnisse, ZEVrail Glasers Annalen, 128, 2004, 11–12, 526–537.

6 Controls and Identification

Stefan Jakubek and Martin Kozek

CONTENTS

6.1 State-Space Models 160
6.1.1 Introduction 160
6.1.2 State-Space Representation for Linear Dynamic Systems 161
6.1.3 Transformation of State-Space Systems 164
6.1.4 Transfer Function Matrices 165
6.1.5 Stability of LTI Systems 166
6.1.6 Controllability and Observability 166
6.1.6.1 Controllability 166
6.1.6.2 Observability 167
6.2 State Feedback Control Design 168
6.2.1 State Feedback Control by Pole Placement 168
6.2.1.1 Pole Placement by Comparison of Coefficients 168
6.2.1.2 Pole Placement by Eigenstructure Assignment 170
6.2.1.3 State Feedback with Integration of the Control Error 171
6.2.2 State Estimation by Observers 173
6.2.2.1 Observer Design by Pole Assignment 174
6.3 Optimal Control 178
6.3.1 Riccati Equation 178
6.3.2 Riccati Equation for $T \rightarrow \infty$ 178
6.3.3 Q–R Selection 180
6.3.3.1 Basics 180
6.3.3.2 State Weighting versus Control Weighting 180
6.3.3.3 Stability 180
6.3.3.4 Modal Weighting 182
6.3.3.5 Weighting of State Derivatives 182
6.4 Identification—Experimental Modeling 183
6.4.1 Introduction 183
6.4.2 Fundamentals 185
6.4.2.1 Reality, System, and Model 185
6.4.3 Model Validation 185
6.4.4 Basics of Discrete-Time Signals and Systems 186
6.4.4.1 Pulse Transfer Function 187
6.4.5 Basics of Probability and Stochastics 187
6.4.5.1 Probability of an Event 187
6.4.5.2 Definition of a Random Variable 187
6.4.5.3 Covariance Matrix 189
6.4.5.4 Auto- and Cross-Correlation Functions 190
6.5 Identification Methods 191
6.5.1 Input Signals and Persistent Excitation 191
6.5.2 Nonparametric Modeling 191
6.5.2.1 Step Response 191

6.5.2.2 Correlation Analysis 191
6.5.2.3 Spectral Analysis 192
6.5.3 Parametric Modeling 192
6.5.3.1 Problem Formulation 192
6.5.3.2 Least-Squares Algorithm 194
6.5.3.3 Recursive Least-Squares Algorithm 196
6.5.3.4 Algorithms for the ARMAX Model 197
6.5.3.5 RLS for Time-Varying Systems 198
6.5.3.6 Prediction Error Method 199
6.5.3.7 Identification of a State-Space Model 199
6.5.4 Black-Box versus Gray-Box Models 199
6.5.5 Nonlinear Modeling 199
6.5.5.1 Models that are Linear in the Parameters 200
6.5.5.2 Local Linear Models 200
6.5.5.3 Artificial Neural Networks 200
6.5.6 Stepwise Identification Procedure 201
6.5.6.1 Gathering Information 201
6.5.6.2 Pre-Processing of Data 201
6.5.6.3 Nonparametric Models 201
6.5.6.4 Parametric models 201
6.5.6.5 Model Validation 202
6.6 Nonlinear Control Design 204
6.6.1 Linearization 204
6.6.1.1 Linearization about an Equilibrium Point 204
6.6.2 Feedback Linearization 205
6.6.2.1 Zero Dynamics—Output Zeroing 206
6.6.3 Robust Control 209
6.7 State Estimation Revisited 211
6.7.1 Kalman Filter 211
6.7.1.1 Motivation 211
6.7.1.2 Analysis of Covariance 212
6.7.1.3 Derivation of the Filter Equations 213
6.7.1.4 Properties of the Kalman Filter 214
6.7.1.5 Stochastic Separation Principle 215
6.7.1.6 Discrete-Time Kalman Filter 215
6.7.1.7 Stability Problems and Divergence 216
6.7.1.8 Related Filters for Nonlinear Systems 216
6.7.2 Sliding Mode Observer 216
6.7.2.1 Motivation 216
6.7.2.2 Principle of the SMO 217
6.7.2.3 Structure of a SMO 217
References 219

6.1 STATE-SPACE MODELS

6.1.1 Introduction

The state-space representation of a dynamic system is given by a system of first-order differential equations. Instead of a linear or nonlinear differential equation of order n as an input/output representation, a system of n linear or nonlinear coupled differential equations of first order is obtained.

6.1.2 State-Space Representation for Linear Dynamic Systems

The state-space representation for linear dynamic systems generally consists of a state differential equation

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{E}\boldsymbol{z}; \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{6.1}$$

and of the algebraic output equation

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{F}\boldsymbol{z}. \tag{6.2}$$

The individual variables in 6.1 and 6.2 denote the following quantities:

$\boldsymbol{x}$: $(n \times 1)$—state vector
$\boldsymbol{u}$: $(r \times 1)$—vector of inputs
$\boldsymbol{y}$: $(m \times 1)$—vector of measurements
$\boldsymbol{z}$: $(q \times 1)$—disturbance vector
$\boldsymbol{A}$: $(n \times n)$—system matrix
$\boldsymbol{B}$: $(n \times r)$—input matrix
$\boldsymbol{C}$: $(m \times n)$—output matrix
$\boldsymbol{D}$: $(m \times r)$—direct feed-through matrix
$\boldsymbol{E}$: $(n \times q)$—system disturbance matrix
$\boldsymbol{F}$: $(m \times q)$—output disturbance matrix

The disturbances can be further distinguished by their nature:

- *Process faults*: They affect the state differential equation and thus have a direct influence on the state vector. They are described by the matrix $\boldsymbol{E}$. Typical examples would be unknown wind forces on a vehicle, increased bearing friction due to wear, increased drag due to low tire pressure or excitation by a road profile.
- *Measurement faults*: They only affect the output equation and are thus described by the matrix $\boldsymbol{F}$. In practice, they typically occur as sensor noise, drift, offset, or sensor breakdown.

In Figure 6.1, Equations 6.1 and 6.2 are depicted as a block diagram.

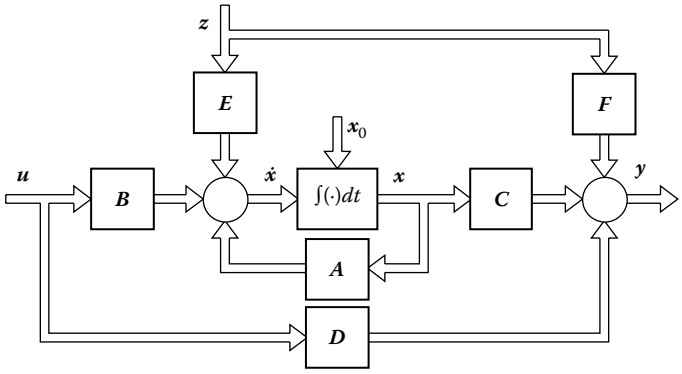


FIGURE 6.1 State-space representation of a multiple-input–multiple-output (MIMO) system.

Example 6.1

Linear single-track vehicle model for simulation and control design.

The single-track model is a simple linear vehicle model commonly used in the analysis and control design of lateral and yaw dynamics. The wheels of the each axle are considered lumped together in the center of the vehicle. The roll, pitch, and heave motions are neglected. In Figure 6.2, the single-track model is illustrated.

Its main variables and geometric parameters are

δ	Steering angle
β	Chassis sideslip angle at center of gravity (CG)
v_x, v_y	Components of velocity vector at CG in x-/y-directions
v_c	Magnitude of velocity vector at CG
a_y	Acceleration at CG in y-direction
v_V, v_H	Velocity vectors at wheel–road contact points
F_{xV}, F_{xH}	Tangential wheel forces
F_{yV}, F_{yH}	Lateral wheel forces
α_V, α_H	Sideslip angles of front and rear wheels, respectively
M	External disturbance torque or control torque
$\dot{\psi}$, $\ddot{\psi}$	Yaw angular velocity and acceleration
l_H, l_V	Distance from rear (front) axle to CG
l	Distance from rear to front axle
m, I_c	Vehicle mass and yaw moment of inertia

The linearized equations of motion are made under the following assumptions:

- Small sideslip and steering angles.
- Inertia effects due to the relative motion of wheels against chassis are neglected.

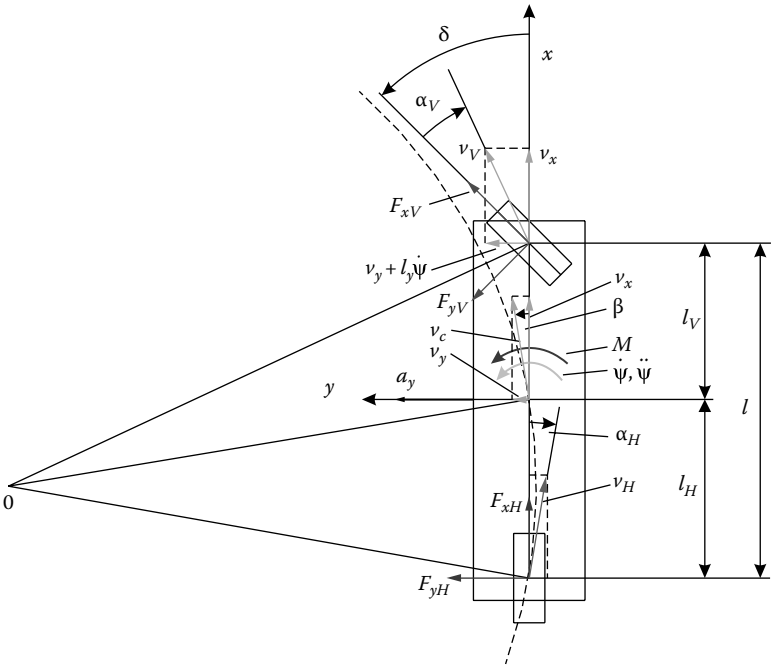


FIGURE 6.2 Single-track model: kinematics, forces, and geometry.

- The travel speed v_x is assumed to remain constant.
- The lateral tire force characteristics are given in linear form: $F_yV = k_V\alpha_V$; $F_yH = k_H\alpha_H$.
- The variability of aerodynamic forces and torques is neglected.
- The effect of different brake forces on individual wheels, for example, during interaction of ESP is modeled by the control torque M.

The equations of motion are given by Equations 6.3 through 6.5, where $l_V^* = l_V$, $l_H^* = -l_H$:

$$\text{Momentum balance in } x\text{-direction:} \quad m(-\dot{\psi}v_y) = \sum_{i=V,H} (F_{xi}\cos\delta_i - F_{yi}\sin\delta_i), \tag{6.3}$$

$$\text{Momentum balance in } y\text{-direction:} \quad m(\dot{v}_y + \dot{\psi}v_x) = \sum_{i=V,H} (F_{xi}\sin\delta_i + F_{yi}\cos\delta_i), \tag{6.4}$$

$$\text{Angular momentum balance:} \quad I_c\ddot{\psi} = \sum_{i=V,H} l_i^*(F_{xi}\sin\delta_i + F_{yi}\cos\delta_i) + M. \tag{6.5}$$

The rear steering wheel angle $\delta_H = 0$ is set to zero in this example. The sideslip angles α_i are determined by (cf. Figure 6.2)

$$\alpha_i = \delta_i - \arctan\left(\frac{v_y + l_i^*\dot{\psi}}{v_x}\right), \quad i = V, H. \tag{6.6}$$

Linearization of the equations of motion yields

$$\dot{v}_y = \dot{\psi}\frac{1}{m}\left(-v_x m - \frac{k_V l_V}{v_x} + \frac{k_H l_H}{v_x}\right) - v_y\frac{1}{m}\left(\frac{k_V}{v_x} + \frac{k_H}{v_x}\right) + \frac{k_V}{m}\delta, \tag{6.7}$$

$$\ddot{\psi} = \dot{\psi}\frac{1}{I_c v_x}(-l_V^2 k_V - l_H^2 k_H) + v_y\frac{1}{I_c v_x}(l_H k_H - l_V k_V) + M\frac{1}{I_c} + \delta\frac{l_V k_V}{I_c}. \tag{6.8}$$

The state vector is chosen as

$$\boldsymbol{x} = \begin{bmatrix} v_y \\ \dot{\psi} \end{bmatrix} \tag{6.9}$$

which results in the following state differential equation for the single-track model:

$$\begin{bmatrix} \dot{v}_y \\ \ddot{\psi} \end{bmatrix} = \begin{bmatrix} \dfrac{-(k_H + k_V)}{m v_x} & \dfrac{(-v_x^2 m - k_V l_V + k_H l_H)}{m v_x} \\ \dfrac{(l_H k_H - l_V k_V)}{I_c v_{\boldsymbol{x}}} & \dfrac{(-l_V^2 kV - l_H^2 k_H)}{I_c v_x} \end{bmatrix} \begin{bmatrix} v_y \\ \dot{\psi} \end{bmatrix} + \begin{bmatrix} \dfrac{k_V}{m} & 0 \\ \dfrac{l_V k_V}{I_c} & \dfrac{1}{I_c} \end{bmatrix} \begin{bmatrix} \delta \\ M \end{bmatrix}. \tag{6.10}$$

In contrast to the state differential equation (6.1), the output equation (6.2) rather is determined by quantities of interest by the user than by physical equations. For vehicle control concepts like ESP, the sideslip angle β at the CG is of particular interest. From Figure 6.2, it is apparent that $\beta = v_y/v_x$. This results in the following output equation:

$$y(t) = \begin{bmatrix} \dfrac{1}{v_x} & 0 \end{bmatrix} \boldsymbol{x}(t) + [\boldsymbol{0}]u(t). \tag{6.11}$$

Given the following parameters:

m	1300 [kg]	I_c	1300 [kg m^2]
l_V	1.9 [m]	l_H	0.8 [m]
v_x	19 [m/s]	k_H	1.4×10^5 [N]
k_V	5.9×10^4 [N]	g	9.81 [m/s^2]

The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ result in

$$\boldsymbol{A} = \begin{bmatrix} -8.0713 & -19.0000 \\ 0 & -12.2683 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 45.43 & 0 \\ 86.33 & 7.6923 \times 10^{-4} \end{bmatrix}, \quad \boldsymbol{C} = [0.0526 \quad 0]. \tag{6.12}$$

6.1.3 Transformation of State-Space Systems

For a given physical system, the choice of the state vector is not unique. Generally, there are infinitely many state representations of one and the same system. The transition from one state representation with state vector $\boldsymbol{x}$ to a different one with state vector $\hat{x}$ can be described by a non-singular $n \times n$ transformation matrix $\boldsymbol{T}$:

$$\boldsymbol{x} = \boldsymbol{T}\hat{\boldsymbol{x}} \quad \text{or} \quad \hat{\boldsymbol{x}} = \boldsymbol{T}^{-1}\boldsymbol{x}, \tag{6.13}$$

Inserting $\boldsymbol{x} = \boldsymbol{T}\hat{\boldsymbol{x}}$ into the state differential equation (6.1) and output equation (6.2), one obtains

$$\boldsymbol{T}\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\boldsymbol{T}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{E}\boldsymbol{z},$$

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{T}\hat{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{F}\boldsymbol{z},$$

where the state differential equation can be transformed to

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}\,\hat{\boldsymbol{x}} + \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{u} + \boldsymbol{T}^{-1}\boldsymbol{E}\boldsymbol{z}.$$

Using the relations

$$\hat{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}, \quad \hat{\boldsymbol{B}} = \boldsymbol{T}^{-1}\boldsymbol{B}, \quad \hat{\boldsymbol{E}} = \boldsymbol{T}^{-1}\boldsymbol{E}, \quad \hat{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{T}, \quad \hat{\boldsymbol{D}} = \boldsymbol{D}, \quad \hat{\boldsymbol{F}} = \boldsymbol{F}, \tag{6.14}$$

the transformed system is obtained:

$$\dot{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{A}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{B}}\boldsymbol{u} + \hat{\boldsymbol{E}}\boldsymbol{z}; \hat{\boldsymbol{x}}(t_0) = \hat{\boldsymbol{x}}_0, \tag{6.15}$$

$$\boldsymbol{y} = \hat{\boldsymbol{C}}\hat{\boldsymbol{x}} + \hat{\boldsymbol{D}}\boldsymbol{u} + \hat{\boldsymbol{F}}\boldsymbol{z}. \tag{6.16}$$

Important properties of state transformations:

1. The determinant of a matrix $\boldsymbol{A}$ is invariant against the described transformation, that is, $|\boldsymbol{A}| = |\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}|$.
2. The eigenvalues of $\boldsymbol{A}$ are also invariant against the given transformation, that is, $|s\boldsymbol{I} - \boldsymbol{A}| = |s\boldsymbol{I} - \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}|$.

6.1.4 Transfer Function Matrices

State-space representations (6.1) and (6.2) can be uniquely transformed to transfer functions. The inverse transform, for example, a conversion from transfer functions to a state-space system is also possible; the result, however, is not unique (cf. 6.3). Setting $\boldsymbol{x}_0 = \boldsymbol{0}$, the Laplace transforms of the state differential equation and the output equation are

$$s\boldsymbol{X}(s) = \boldsymbol{A}\boldsymbol{X}(s) + \boldsymbol{B}\boldsymbol{U}(s) + \boldsymbol{E}\boldsymbol{Z}(s), \tag{6.17}$$

$$\boldsymbol{Y}(s) = \boldsymbol{C}\boldsymbol{X}(s) + \boldsymbol{D}\boldsymbol{U}(s) + \boldsymbol{F}\boldsymbol{Z}(s). \tag{6.18}$$

Solving (6.18) for $\boldsymbol{X}(s)$ yields $\boldsymbol{X}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B}\boldsymbol{U}(s) + (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{E}\boldsymbol{Z}(s)$. The matrix $\boldsymbol{\Phi}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}$ is called the transition matrix. Inserting $\boldsymbol{X}(s)$ into 6.18 leads to

$$\boldsymbol{Y}(s) = [\boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}]\boldsymbol{U}(s) + [\boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{E} + \boldsymbol{F}]\boldsymbol{Z}(s). \tag{6.19}$$

The $(m \times r)$-transfer function matrix $\boldsymbol{G}_{SU}(s) = \boldsymbol{Y}(s)/\boldsymbol{U}(s)$ and the $(m \times q)$-transfer function matrix $\boldsymbol{G}_{SZ}(s) = \boldsymbol{Y}(s)/\boldsymbol{Z}(s)$ are therefore given by

$$\boldsymbol{G}_{SU}(s) = \boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}, \tag{6.20}$$

$$\boldsymbol{G}_{SZ}(s) = \boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{E} + \boldsymbol{F}. \tag{6.21}$$

Example 6.2

The following state-space system is considered:

$$\boldsymbol{A} = \begin{bmatrix} -5 & 2 \\ -6 & 2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \boldsymbol{B} = [1 \quad -1], \quad \boldsymbol{D} = 0, \quad \boldsymbol{E} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{F} = 0.$$

The transition matrix is

$$\boldsymbol{\Phi}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1} = \begin{bmatrix} s+5 & -2 \\ 6 & s-2 \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{s-2}{(s+1)(s+2)} & \dfrac{2}{(s+1)(s+2)} \\ \dfrac{-6}{(s+1)(s+2)} & \dfrac{s+5}{(s+1)(s+2)} \end{bmatrix}.$$

The (scalar) transfer function $G_{SU}(s)$ is obtained from (6.20):

$$G_{SU}(s) = [1 \quad -1] \begin{bmatrix} \dfrac{s-2}{(s+1)(s+2)} & \dfrac{2}{(s+1)(s+2)} \\ \dfrac{-6}{(s+1)(s+2)} & \dfrac{s+5}{(s+1)(s+2)} \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{2}{(s+1)(s+2)}.$$

The (scalar) transfer function $G_{SZ}(s)$ is obtained from (6.21)

$$G_{sz}(s) = [1 \quad -1] \begin{bmatrix} \dfrac{s-2}{(s+1)(s+2)} & \dfrac{2}{(s+1)(s+2)} \\ \dfrac{-6}{(s+1)(s+2)} & \dfrac{s+5}{(s+1)(s+2)} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{s+4}{(s+1)(s+2)}.$$

6.1.5 Stability of LTI Systems

The stability of a state-space system is determined by the homogeneous part of its state differential equation:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}. \tag{6.22}$$

Stability is determined by means of the eigenvalues of the matrix $\boldsymbol{A}$. They are the roots of the characteristic polynomial $P(\lambda)$ given by

$$P(\lambda) = |\lambda \boldsymbol{I} - \boldsymbol{A}| = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_2\lambda^2 + a_1\lambda + a_0 \tag{6.23}$$

Theorem 6.1

*The LTI system (6.22) is **asymptotically stable**, if the real parts of all eigenvalues of **A** are negative.*

Theorem 6.2

*The LTI system (6.22) is **marginally stable**, if the real parts of all eigenvalues of **A** are negative, except for an imaginary couple or an eigenvalue equal to zero both of which with a multiplicity of one.*

Theorem 6.3

*The LTI system (6.22) is **unstable**, if the real part of an eigenvalue of **A** is either positive or if eigenvalues of **A** with a multiplicity greater than one are located on the imaginary axis.*

6.1.6 Controllability and Observability

6.1.6.1 Controllability

Definition 6.1

The LTI system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}; \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{6.24}$$

is fully state controllable, if there exists an unconstrained control input $\boldsymbol{u}(t)$ that transfers any initial state $\boldsymbol{x}(t_0)$ to any arbitrary state $\boldsymbol{x}(t_1)$ within a finite time interval $t_0 \le t \le t_1$.

Remarks:

- The earlier definition also implies that every component of the state vector $\boldsymbol{x}$ is directly or indirectly influenced by the input $\boldsymbol{u}$.
- In the earlier definition for controllability, it is not required that $\boldsymbol{x}(t) = \boldsymbol{x}(t_1)$ for $t > t_1$.

Full controllability only depends on the system matrix $\boldsymbol{A}$ and on the input matrix $\boldsymbol{B}$ and can be expressed in terms of the controllability matrix $\boldsymbol{Q}_c$.

Definition 6.2

The system (6.24) is fully state controllable if the ($n \times nr$)-controllability matrix $\boldsymbol{Q}_c$ has full rank n:

$$rank(\boldsymbol{Q}_c) = rank([\boldsymbol{B}, \boldsymbol{AB}, \boldsymbol{A}^2\boldsymbol{B}, \ldots, \boldsymbol{A}^{n-1}\boldsymbol{B}]) = n. \tag{6.25}$$

Example 6.3

The following two systems are considered:

$$\text{Case 1: } \dot{\boldsymbol{x}} = \begin{bmatrix} -5 & 2 \\ -6 & 2 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 2 \\ 2 \end{bmatrix}u \qquad \text{Case 2: } \dot{\boldsymbol{x}} = \begin{bmatrix} -5 & 2 \\ -6 & 2 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 2 \\ 4 \end{bmatrix}u.$$

In case 1 the controllability matrix has rank two which indicates full controllability, whereas in case 2 rank ($\boldsymbol{Q}_c$) = 1 holds. This indicates that there exists an uncontrollable subspace, that is, a domain of the state space that cannot be reached by the control input $u(T)$.

6.1.6.2 Observability

Definition 6.3

The LTI system

$$\dot{\boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{Bu}; \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{6.26}$$

$$\boldsymbol{y} = \boldsymbol{Cx} + \boldsymbol{Du} \tag{6.27}$$

is fully state observable if any initial state $\boldsymbol{x}_0$ can be determined from observation of the input $\boldsymbol{u}(t)$ and the output $\boldsymbol{y}(t)$ within a finite time $t_0 \leq t \leq t_1$.

Remark:
The given definition also implies that every component of the state vector $\boldsymbol{x}$ influences the outputs $\boldsymbol{y}$.

Full state observability only depends on the system matrix $\boldsymbol{A}$ and the output matrix $\boldsymbol{C}$ and can be expressed in terms of the observability matrix $\boldsymbol{Q}_0$.

Theorem 6.4

The system (6.26) and (6.27) is fully state observable if the ($mn \times n$)-observability matrix has full rank n:

$$rank(\boldsymbol{Q}_B) = rank\left(\begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{n-1} \end{bmatrix}\right) = n. \tag{6.28}$$

6.2 STATE FEEDBACK CONTROL DESIGN

6.2.1 State Feedback Control by Pole Placement

Figure 6.3 shows a block diagram of a single-input–single-output system that is controlled by a state feedback controller. The system has the following state-space representation:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u + \boldsymbol{E}z; \quad \boldsymbol{x}_0 = 0, \tag{6.29}$$

$$y = \boldsymbol{c}^T\boldsymbol{x} + du. \tag{6.30}$$

The control law of a state feedback controller is given as

$$u = K_w w - \boldsymbol{k}^T\boldsymbol{x}. \tag{6.31}$$

Full state feedback design commonly relies on pole-placement techniques. If the plant (6.29) is fully state controllable (cf. Section 6.1.6), then the n components of the feedback gain $\boldsymbol{k}^T = [k_1\ k_2\ \ldots\ k_n]$ can be determined uniquely by choosing n eigenvalues of the closed loop. The gain K_w is used to guarantee a closed-loop steady-state gain of one.

By inserting the control law (6.31) in (6.29) and (6.30), the state-space representation of the closed loop is obtained:

$$\dot{\boldsymbol{x}} = (\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T)\boldsymbol{x} + \boldsymbol{b}K_w w + \boldsymbol{E}z, \tag{6.32}$$

$$y = (c^T - d\boldsymbol{k}^T)\boldsymbol{x} + dK_w w + \boldsymbol{f}^T z. \tag{6.33}$$

6.2.1.1 Pole Placement by Comparison of Coefficients

An intuitive approach to determine the state feedback gain $\boldsymbol{k}^T$ is by comparing the coefficients of the characteristic polynomial of the closed-loop system to those of a desired polynomial. From 6.32, the characteristic polynomial of the closed loop results in

$$P_k(s, \boldsymbol{k}^T) = det[s\boldsymbol{I} - (\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T)]. \tag{6.34}$$

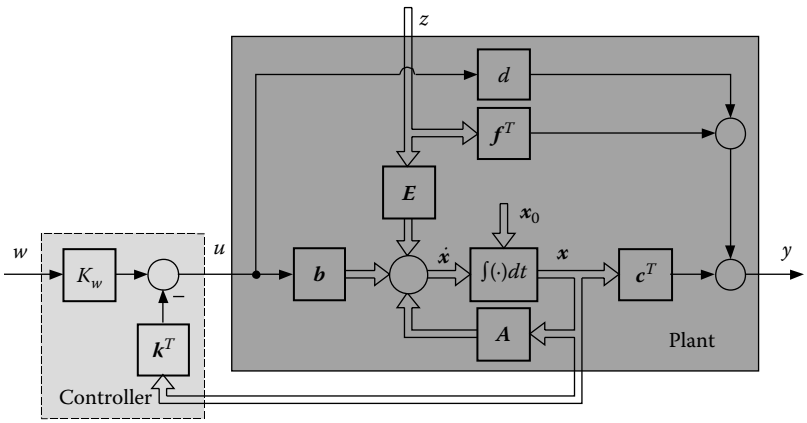


FIGURE 6.3 State feedback control.

The coefficients (6.34) are compared to a desired polynomial $P(s)$, which results in n equations for the n components of $\boldsymbol{k}^T$:

$$P(s) = (s-\lambda_1)(s-\lambda_2)\ \ldots\ (s-\lambda_n). \tag{6.35}$$

The roots λ_i of $P(s)$ are typically determined from dynamic specifications of the closed-loop step response like maximum overshoot e_m, rise time T_r, or settling time T_s.

6.2.1.1.1 Determination of the Gain K_w

The gain K_w is determined from the requirement that the steady-state gain of the closed loop is equal to one so that the output y correctly follows the reference input w.

Setting $\dot{\boldsymbol{x}} = \boldsymbol{0}$ and $\boldsymbol{z} = \boldsymbol{0}$ in 6.32 leads to

$$\boldsymbol{x}_\infty = (\boldsymbol{b}\boldsymbol{k}^T - \boldsymbol{A})^{-1}\boldsymbol{b}K_w w_\infty, \tag{6.36}$$

where ∞ indicates the steady state. Inserting 6.36 into 6.33 yields

$$y_\infty = (\boldsymbol{c}^T - d\boldsymbol{k}^T)\boldsymbol{x}_\infty + dK_w w_\infty = [(\boldsymbol{c}^T - d\boldsymbol{k}^T)(\boldsymbol{b}\boldsymbol{k}^T - \boldsymbol{A})^{-1}\boldsymbol{b} + d]\boldsymbol{K}_w w_\infty.$$

Since $y_\infty = w_\infty$ should hold, K_w must satisfy

$$K_w = [(\boldsymbol{c}^T - d\boldsymbol{k}^T)(\boldsymbol{b}\boldsymbol{k}^T - \boldsymbol{A})^{-1}\boldsymbol{b} + d]^{-1}. \tag{6.37}$$

Example 6.4

The single-track model from Example 6.1 is considered.

As a control input, the yawing torque M is chosen. In reality, M is applied by selective braking of single wheels. For the closed loop, a natural frequency of $\omega_n = 18$ and a damping constant of $\zeta = 0.8$ are chosen. This corresponds to an eigenvalue pair $p_{1,2} = -14.4 \pm 10.8i$ and a desired polynomial P_{des} of

$$P_{des} = s^2 + 28.8s + 324. \tag{6.38}$$

For the characteristic polynomial of the closed loop, one obtains

$$P(s) = s^2 + s(20.3396 + 7.692\times10^{-4}k_2) + (99.0208 + 0.0062k_2 - 0.0146k_1), \tag{6.39}$$

which yields the following comparison of coefficients:

$$\begin{aligned} s^2&: \quad 1 = 1 \\ s^1&: \quad 20.3396 + 7.692\times10^{-4}k_2 = 28.8 \\ s^0&: \quad 99.0208 + 0.0062k_2 - 0.0146k_1 = 324 \end{aligned}$$

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = 10^4 \times \begin{bmatrix} -1.0721 \\ 1.0999 \end{bmatrix}. \tag{6.40}$$

The gain K_w is computed from 6.37:

$$K_w = -4.2120\times10^5. \tag{6.41}$$

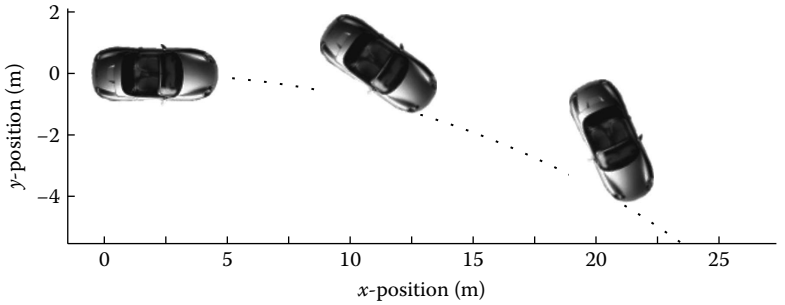


FIGURE 6.4 C/G trajectory without state feedback control.

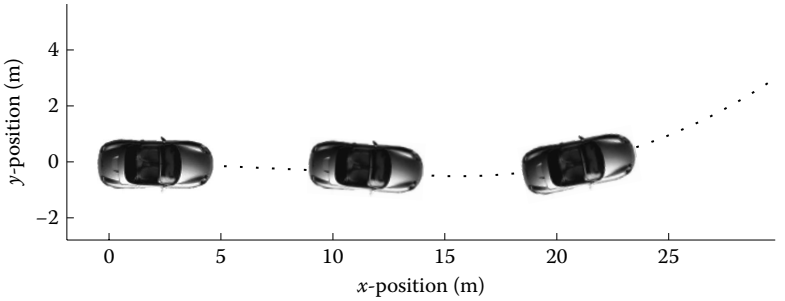


FIGURE 6.5 C/G trajectory with state feedback control.

Figures 6.4 and 6.5 show the C/G trajectories without and with the controller just designed. As initial conditions, $\upsilon_y = 0\frac{m}{s^2}$ and $\dot{\psi} = -2\frac{1}{s}$ were chosen. Additionally, after 1.5 s the driver initiates a left curve by a proper choice of δ. It is apparent that due to the high initial yaw angular velocity the car is oversteering clockwise. Figure 6.4 shows that this undesirable state is not ended without controller interaction, while in Figure 6.5 the car perfectly follows the driver's intention and swerves into a left turn.

6.2.1.2 Pole Placement by Eigenstructure Assignment

Pole placement by comparison of coefficients only delivers unique results for single-input systems. An alternative approach that can also be easily implemented in computer programs is direct eigenstructure assignment.

This approach directly looks at the eigenvalue equation of the system matrix $\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T$ of the closed loop (6.32):

$$(\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T)\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i. \tag{6.42}$$

Rearranging this equation leads to

$$(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{u}_i = \boldsymbol{b}\boldsymbol{k}^T\boldsymbol{u}_i \Rightarrow \boldsymbol{u}_i \underbrace{(\boldsymbol{A} - \lambda_i \boldsymbol{I})^{-1}\boldsymbol{b}}_{\boldsymbol{S}(\lambda_i)} \underbrace{\boldsymbol{k}^T\boldsymbol{u}_i}_{\boldsymbol{z}_i}. \tag{6.43}$$

After choosing an eigenvalue λ_i, the associated eigenvector $\boldsymbol{u}_i$ is obtained from 6.43. For single-input systems, $\boldsymbol{z}_i$ is scalar and can be chosen $z_i = 1$ without loss of generality. This yields

$$\boldsymbol{u}_i = \boldsymbol{S}(\lambda_i). \tag{6.44}$$

With $\boldsymbol{u}_i$ from 6.44 and $z_i = 1$, one further obtains *one* equation for the determination of the feedback gain $\boldsymbol{k}^T$:

$$\boldsymbol{k}^T\boldsymbol{u}_i = 1. \tag{6.45}$$

If n eigenvalues $\lambda_1 \ldots \lambda_n$ are chosen in the manner described earlier, a complete matrix equation for the actual determination of $\boldsymbol{k}^T$ is created by collecting all equations (6.45):

$$\boldsymbol{k}^T[\boldsymbol{u}_1 \quad \ldots \quad \boldsymbol{u}_n] = [1 \quad \ldots \quad 1] \Rightarrow \boldsymbol{k}^T = [1 \quad \ldots \quad 1][\boldsymbol{u}_1 \quad \ldots \quad \boldsymbol{u}_n]^{-1}. \tag{6.46}$$

- For complex–conjugate couples, 6.43 is separated into real and imaginary parts and thus delivers *two* equations for the determination of the feedback gain.
- It is not possible to assign eigenvalues of the open-loop system to the closed loop since $\boldsymbol{S}(\lambda)$ would not exist in this case.
- The inverse in 6.46 can only be computed if the eigenvector matrix $[\boldsymbol{u}_1 \ldots \boldsymbol{u}_n]$ is non-singular. The conditioning of this matrix can be influenced implicitly by the choice of eigenvalues.
- In the multi-input case, $\boldsymbol{z}_i$ in 6.43 has as many elements as there are control inputs. This gives additional design freedom for the eigenvector $\boldsymbol{u}_i$ even after the eigenvalue λ_i has been chosen. This freedom is mostly used to ensure proper conditioning of the eigenvector matrix. An introduction to related methods is given in [12].

6.2.1.3 State Feedback with Integration of the Control Error

A common method to reject unknown disturbances is by additional integration of the control error. For a single-input system with one control variable, the control law now reads

$$u = -\boldsymbol{k}^T\boldsymbol{x} + K\int_0^t (w - y)dt \tag{6.47}$$

and the system representation is given by

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u + \boldsymbol{E}\boldsymbol{z}, \tag{6.48}$$

$$y = \boldsymbol{c}^T\boldsymbol{x} + du + \boldsymbol{f}^T\boldsymbol{z}. \tag{6.49}$$

In order to integrate 6.47 in the state feedback concept, an additional state variable $p(t)$ is defined:

$$p(t) = \int_0^t (y(\tau) - w(\tau))d\tau. \tag{6.50}$$

Differentiating $p(t)$ and inserting the output equation (6.49) yields

$$\dot{p}(t) = y(t) - w(t) = \boldsymbol{c}^T\boldsymbol{x} + du + \boldsymbol{f}^T\boldsymbol{z} - w. \tag{6.51}$$

Next, 6.48 and 6.51 are combined to one augmented state-space system:

$$\begin{bmatrix} \dot{\boldsymbol{x}} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ c^T & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ p \end{bmatrix} + \begin{bmatrix} \boldsymbol{b} \\ d \end{bmatrix}u + \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ f^T & -1 \end{bmatrix}\begin{bmatrix} z \\ w \end{bmatrix}. \tag{6.52}$$

Given the augmented state vector $\hat{\boldsymbol{x}} = [x^T \quad p]^T$, the control law (6.47) now reads

$$u(t) = -\hat{\boldsymbol{k}}^T\hat{\boldsymbol{x}}(t) \quad \text{with}\, \hat{\boldsymbol{k}}^T = [\boldsymbol{k}^T \quad K]. \tag{6.53}$$

The augmented feedback gain vector $\hat{\boldsymbol{k}}^T$ can now be determined with the augmented system (6.52) using the techniques from Sections 6.2.1.1 and 6.2.1.2. Again, full state feedback requires that the system (6.53) is fully state controllable. If the original system (6.48) is fully state controllable, this can be ensured if additionally the matrix

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ c^T & d \end{bmatrix} \tag{6.54}$$

has full rank. Inserting 6.53 in 6.52 gives the state-space representation of the closed loop:

$$\begin{bmatrix} \dot{\boldsymbol{x}} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{b}\boldsymbol{k}^T & -K\boldsymbol{b} \\ \boldsymbol{c}^T & -d\boldsymbol{k}^T & -dK \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ p \end{bmatrix} + \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ \boldsymbol{f}^T & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{z} \\ w \end{bmatrix}. \tag{6.55}$$

The transfer function matrices of the closed-loop system are

$$\boldsymbol{z} = \boldsymbol{0}: \quad G_w(s) = [(c^T - d\boldsymbol{k}^T) - dK]\Phi(s) \begin{bmatrix} \boldsymbol{0} \\ -1 \end{bmatrix}, \tag{6.56}$$

$$w = 0: \quad \boldsymbol{G}_z(s) = [(\boldsymbol{c}^T - d\boldsymbol{k}^T) - dK]\Phi(s) \begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{f}^T \end{bmatrix} + \boldsymbol{f}^T, \tag{6.57}$$

where

$$\Phi(s) = \begin{bmatrix} (s\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{b}\boldsymbol{k}^T) & Kb \\ (d\boldsymbol{k}^T - \boldsymbol{c}^T) & s + dK \end{bmatrix}^{-1}.$$

Example 6.5

The following system is considered:

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0, & 25 \\ 0 & \end{bmatrix} u + \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \boldsymbol{z},$$

$$y = [0 \quad 1]\boldsymbol{x}.$$

A state feedback controller with integration of the control error has to be designed such that the poles of the closed loop are located at $s_{1,2} = -2 \pm j_2$ and $s_3 = -10$.

First, the criteria for full state controllability of the augmented system (6.52) have to be checked:

$$rank([\,\boldsymbol{b} \quad \boldsymbol{A}\boldsymbol{b}\,]) = 2 \text{ and } rank\left(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{c}^T & 0 \end{bmatrix}\right) = 3$$

The characteristic polynomial of the closed-loop system is

$$P_k(s, \boldsymbol{k}^T, K) = det[s\boldsymbol{I} - \hat{\boldsymbol{A}} + \hat{\boldsymbol{b}}\,[\boldsymbol{k}^T \quad K]] = \begin{vmatrix} s+1+0.25k_1 & -1+0.25k_2 & 0.25K \\ -2 & s+2 & 0 \\ 0 & -1 & s \end{vmatrix} =$$

$$= s^3 + (3+0.25k_1)s^2 + (0.5k_1 + 0.5k_2)s + 0.5K$$

and the desired polynomial $P(s)$ is

$$P(s) = (s^2 + 4s + 8)(s + 10) = s^3 + 14s^2 + 48s + 80.$$

A comparison of coefficients yields

$$3 + 0.25k_1 = 14 \quad k_1 = 44$$

$$0.5k_1 + 0.5k_2 = 48 \Rightarrow k_2 = 52 \Rightarrow u = -[44 \quad 52]\boldsymbol{x} + 160\int(w - y)dt$$

$$0.5K = 80 \quad K = 160.$$

Equations 6.55 through 6.57 deliver the transfer function matrices:

$$G_w(s) = [(\boldsymbol{c}^T - d\boldsymbol{k}^T) - dK]\Phi(s)\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \frac{80}{s^3 + 14s^2 + 48s + 80},$$

$$\boldsymbol{G}_z(s) = [(\boldsymbol{c}^T - d\boldsymbol{k}^T) - dK]\Phi(s)\begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{f}^T \end{bmatrix} = \boldsymbol{f}^T = \begin{bmatrix} \dfrac{-2s}{s^3 + 14s^2 + 48s + 80} & \dfrac{-2s(s+12)}{s^3 + 14s^2 + 48s + 80} \end{bmatrix}.$$

It is apparent that $G_w(s)$ has a steady-state gain of one whereas both elements of $G_Z(s)$ have globally differentiating behavior.

6.2.2 State Estimation by Observers

In the previous sections, it was assumed that all elements of the state vector $\boldsymbol{x}(t)$ are available. In most applications, however, this is not the case. Often, some elements of the state vector are not measurable at all. In these situations, an estimate $\hat{\boldsymbol{x}}$ of the state vector $\boldsymbol{x}$ is computed by the so-called Luenberger state observer. It is assumed that a number of measurements from the system is available that are combined in a measurement vector $m(t)$. The dynamic system is thus described by

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \tag{6.58}$$

$$\boldsymbol{m} = \boldsymbol{C}^*\boldsymbol{x} + \boldsymbol{D}^*\boldsymbol{u}. \tag{6.59}$$

A full Luenberger observer is given by

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{H}(m - \boldsymbol{C}^*\hat{\boldsymbol{x}} - \boldsymbol{D}^*\boldsymbol{u}), \tag{6.60}$$

where

$\hat{\boldsymbol{x}}$ is the estimate of the state vector
$\boldsymbol{H}$ is the observer gain matrix

It is apparent that the Luenberger observer is by itself a state-space system with $\hat{\boldsymbol{x}}$ as the state variable and $u(t)$ and $\boldsymbol{m}(t)$ as inputs. The reconstruction error is given as $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{x}}$. The dynamic behavior of $\tilde{\boldsymbol{x}}(t)$ is obtained by subtraction of 6.60 from 6.58:

$$\dot{\tilde{\boldsymbol{x}}} = (\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*)\tilde{\boldsymbol{x}}. \tag{6.61}$$

If the system (6.58) and (6.59) is fully state observable, the poles of the observer system matrix $\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*$ can be assigned arbitrarily by proper choice of the observer gain matrix $\boldsymbol{H}$ (with exception of the poles of $\boldsymbol{A}$).

6.2.2.1 Observer Design by Pole Assignment

6.2.2.1.1 Observer Design by Comparison of Coefficients

If the measurement vector $\boldsymbol{m}$ contains only one element, the observer gain matrix can be determined by comparison of coefficients. The characteristic polynomial of the system (6.61) is

$$P_B(s, \boldsymbol{H}) = det(s\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{H}\boldsymbol{C}^*). \tag{6.62}$$

The n eigenvalues of $\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*$ are determined by the elements of the $(1 \times n)$ observer gain $\boldsymbol{H}$ by comparison to a desired polynomial:

$$P_B(s) = (s - \lambda_{B1})(s - \lambda_{B2}) \ldots (s - \lambda_{Bn}). \tag{6.63}$$

Observer poles should generally be significantly faster than the poles of the closed loop (see Example 6.6).

6.2.2.1.2 Observer Design by Eigenstructure Assignment

Eigenstructure assignment aims at assigning eigenvalues and eigenvectors to the observer system matrix $\boldsymbol{F} = \boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*$. Here, the following approach is chosen:

$$\boldsymbol{u}_i^T(\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*) = \boldsymbol{u}_i^T\lambda_i. \tag{6.64}$$

In 6.64, $\boldsymbol{u}_i^T$ is a *left eigenvector* of $\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*$. Taking the transpose of 6.64 yields

$$(\boldsymbol{A}^T - \boldsymbol{C}^{*T}\boldsymbol{H}^T)\boldsymbol{u}_i = \boldsymbol{u}_i\lambda_i. \tag{6.65}$$

Formally, 6.65 resembles an eigenvalue equation of the form 6.42. Comparing the associated terms in 6.42 and 6.65 shows that there is a duality between state feedback control and observation. This duality is highlighted in the following table:

Controller	Observer
$\boldsymbol{A}$	$\boldsymbol{A}^T$
$\boldsymbol{B}$	$\boldsymbol{C}^{*T}$
$\boldsymbol{K}$	$\boldsymbol{H}^T$

Using this information, the results from Section 6.2.1.2 can be applied almost directly for observer design.

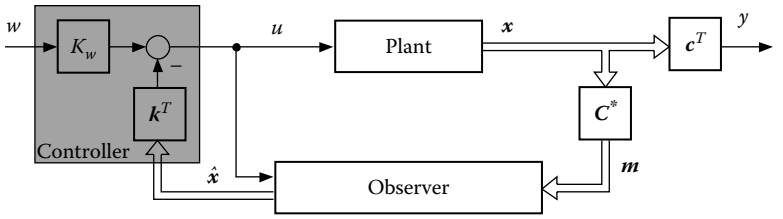


FIGURE 6.6 Luenberger observer in the control loop.

6.2.2.1.3 *Separation Principle*

Figure 6.6 shows the block diagram of a state feedback control loop where instead of the actual state vector $\boldsymbol{x}$ its estimate $\hat{\boldsymbol{x}}$ is used in the control law:

$$u = K_w w - \boldsymbol{k}^T \hat{\boldsymbol{x}}. \tag{6.66}$$

An observer in series with the control law (6.66) is commonly called *compensator.*

One important question to be answered is how the closed-loop system behaves if the control law (6.66) is applied. Since $\hat{\boldsymbol{x}}$ is only an estimate of $\boldsymbol{x}$, it would be justified to cast doubts on the stability of the resulting closed loop.

The answer to this question is given by the separation principle.

Theorem 6.5

Separation principle

If the system under consideration is fully state controllable and fully state observable, the poles of the closed loop (6.32) and the poles of the observer (6.61) can be assigned independently. The poles of the closed loop with control law (6.66) are then the combination of the controller poles and observer poles. The characteristic equation of the system then is

$$det[s\boldsymbol{I} - (\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T)] \cdot det[s\boldsymbol{I} - (\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}^*)] = 0. \tag{6.67}$$

Example 6.6

The following system is considered:

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0.25 \\ 0 \end{bmatrix} u + \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \boldsymbol{z} \quad \text{and} \quad y = \begin{bmatrix} 0 & 1 \end{bmatrix} \boldsymbol{x}.$$

Based on the output $y(t)$, an observer has to be designed with two real poles at –20 ($\lambda_{B1} = \lambda_{B2} = -20$.)

The rank of the observability matrix is two so that full state observability can be achieved with $y(t)$. The observer system matrix is given by

$$\boldsymbol{A} - \boldsymbol{h}\boldsymbol{c}^T = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} - \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} [0 \quad 1] = \begin{bmatrix} -1 & 1 & -h_1 \\ 2 & -2 & -h_2 \end{bmatrix}.$$

The characteristic polynomial of the observer thus is

$$P_B(s, \boldsymbol{h}) = Det(s\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{h}\boldsymbol{c}^T) = \begin{vmatrix} s+1 & h_1 - 1 \\ -2 & s+2+h_2 \end{vmatrix} = s^2 + (h_2 + 3)s + h_2 + 2h_1$$

and with $\lambda_{B1} = \lambda_{B2} = -20$ the desired polynomial is

$$P_B(s) = s^2 + 40s + 400.$$

A comparison of coefficients yields $\boldsymbol{h}^T = [181.5\ 37]$.

The observer equation follows from 6.60:

$$\dot{\hat{x}} = \boldsymbol{A}\hat{x} + \boldsymbol{b}u + \boldsymbol{H}(m - \boldsymbol{C}^*\hat{\boldsymbol{x}}) = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}\hat{\boldsymbol{x}} + \begin{bmatrix} 0,25 \\ 0 \end{bmatrix} u + \begin{bmatrix} 181.5 \\ 37 \end{bmatrix}[0\ \ 1](\boldsymbol{x} - \hat{\boldsymbol{x}}),$$

$$\dot{\hat{\boldsymbol{x}}} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}\hat{\boldsymbol{x}} + \begin{bmatrix} 0 & 25 \\ 0 & \end{bmatrix} u + \begin{bmatrix} 181.5 \\ 37 \end{bmatrix}(x_2 - \hat{x}_2).$$

Example 6.7

The single-track model from Example 6.1 is revisited:

The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are (6.12)

$$\boldsymbol{A} = \begin{bmatrix} -8.0713 & -19.0000 \\ 0 & -12.2683 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 45.43 & 0 \\ 86.33 & 7.6923 \times 10^{-4} \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 0.0526 \\ 0 \end{bmatrix}. \tag{6.68}$$

It is assumed that the yaw angular velocity $\dot{\psi}$ and the lateral acceleration $\dot{v}_y$ are available as measurements. In this case, the measurement vector contains the derivative of a state:

$$\boldsymbol{m} = \begin{bmatrix} x_2 \\ \dot{x}_1 \end{bmatrix} = \begin{bmatrix} \dot{\psi} \\ \dot{v}_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}). \tag{6.69}$$

Using the first row vectors of $\boldsymbol{A}$ and $\boldsymbol{B}$, this leads to

$$\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ a_{11} & a_{12} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ b_{11} & b_{12} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{6.70}$$

and in matrix notation

$$\boldsymbol{m} = \boldsymbol{C}^* x + \boldsymbol{D}^* u. \tag{6.71}$$

In the given example, $\boldsymbol{C}^*$ and $\boldsymbol{D}^*$ result in

$$\boldsymbol{C}^* = \begin{bmatrix} 0 & 1 \\ -8.0713 & -19 \end{bmatrix}; \quad \boldsymbol{D}^* = \begin{bmatrix} 0 & 0 \\ 45.4382 & 0 \end{bmatrix}. \tag{6.72}$$

Since the observer poles should be significantly faster than those of the closed loop, they are chosen as $\lambda_{1,2} = -30, -31$.

In the present example, the number of measurements in $\boldsymbol{m}$ is equal to the number of states so that for the determination of the elements of the gain matrix $\boldsymbol{H}$ the following strategy can be applied.

The closed-loop observer matrix $\boldsymbol{A} - \boldsymbol{HC}^*$ is chosen such that the estimation errors are decoupled, that is

$$\boldsymbol{A} - \boldsymbol{HC}^* = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Solving for $\boldsymbol{H}$ leads to

$$\boldsymbol{H} = \left(\boldsymbol{A} - \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \right) (C^*)^{-1}. \tag{6.73}$$

For regular $n \times n$ matrices, $\boldsymbol{C}^*$ (6.73) yields a unique solution for $\boldsymbol{H}$:

$$\boldsymbol{H} = \begin{bmatrix} -70.6209 & -2.7169 \\ 18.7317 & 0 \end{bmatrix}. \tag{6.74}$$

Figures 6.7 and 6.8 show time histories of the state estimates $\dot{v}_y$ and $\hat{\dot{\psi}}$ compared to the true values. For comparison, the estimates $\hat{v}_{y,slow}$ and $\hat{\dot{\psi}}_{slow}$ from a slower observer with poles $\lambda_{1,slow}$, $\lambda_{2,slow} = 14.5, -15$ are also shown.

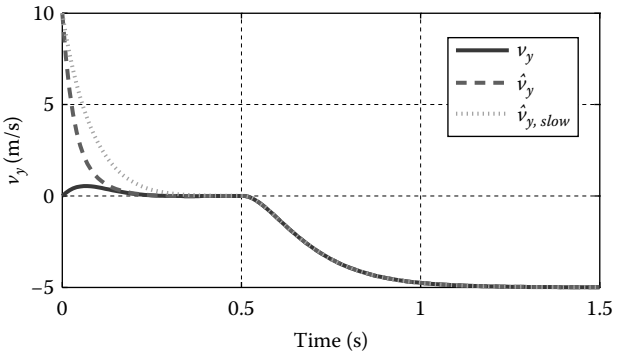


FIGURE 6.7 Time history of v_y, $\hat{v}_y$, and $\hat{v}_{y,slow}$.

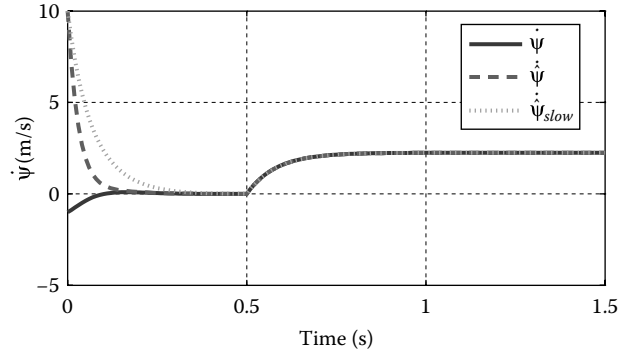


FIGURE 6.8 Time history of $\dot{\psi}$, $\hat{\dot{\psi}}$, and $\hat{\dot{\psi}}_{slow}$.

6.3 OPTIMAL CONTROL

The design of optimal control systems is an important feature of control engineering. Control systems in many different fields are designed on the basis of a performance criterion. Typical examples are

- The way different species move about
- The orientation of plants according to the position of the sun
- Fund management

Control action in the earlier examples aims at maximizing some kind of benefit or minimizing some cost.

This chapter presents the concept of optimal control for linear state-space systems. The obtained control principle is called *linear quadratic regulator* (LQR) *control*.

6.3.1 Riccati Equation

Given the linear state-space representation

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \qquad \boldsymbol{x}(t=0) = \boldsymbol{x}_0. \tag{6.75}$$

The following cost function will be minimized over a time horizon T:

$$V(\boldsymbol{x}(0), \boldsymbol{u}(\cdot)) = \int_0^T \frac{1}{2}(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u})dt + \frac{1}{2}\boldsymbol{x}(T)^T\boldsymbol{G}\boldsymbol{x}(T). \tag{6.76}$$

The weighting matrices $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{G}$ must be symmetric and positive (semi)definite. The advantage of the choice of the cost function (6.76) lies in the fact that its minimization leads to a linear state feedback law of the form $\boldsymbol{u}(t) = \boldsymbol{K}(t)\boldsymbol{x}(t)$.

The minimization of 6.76 by the system 6.75 is accomplished through the following control law:

$$\boldsymbol{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}(t)\boldsymbol{x}(t). \tag{6.77}$$

Here, $\boldsymbol{P}(t)$ is the solution of the *Riccati* equation:

$$-\dot{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{Q} \qquad \text{with } \boldsymbol{P}(T) = \boldsymbol{G}. \tag{6.78}$$

The Ricatti equation is a differential equation for the matrix $\boldsymbol{P}(t)$, which is solved backward in time. Due to its symmetric nature, $\boldsymbol{P}(t)$ is also symmetric.

6.3.2 Riccati Equation for $T \to \infty$

In most practical applications, $T \to \infty$ is chosen for the time horizon. The cost function (6.76) now reads

$$V(\boldsymbol{x}(0), \boldsymbol{u}(\cdot)) = \int_0^\infty \frac{1}{2}(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u})dt \tag{6.79}$$

whereby the terminal cost term $\frac{1}{2}\boldsymbol{x}(T)^T\boldsymbol{G}\boldsymbol{x}(T)$ has been omitted. As opposed to 6.76, the cost function (6.79) does not always have a minimum.

Theorem 6.6 (Stabilizability)

A linear state-space system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \qquad \boldsymbol{x}(t=0) = \boldsymbol{x}_0$$

is stabilizable if all noncontrollable eigenmodes are asymptotically stable, that is, if the real parts of their associated eigenvalues are negative.

For $T \to \infty$ and for positive definite weighting matrices $\boldsymbol{Q}$, $\boldsymbol{R}$ the cost function (6.79) is bounded only if the underlying system is stabilizable. In this case, the Riccati equation converges toward a stationary $\boldsymbol{P}$ which can be obtained from 6.78 by setting $\dot{\boldsymbol{P}}$ to zero:

$$\boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{0}. \tag{6.80}$$

Equation 6.96 is called *Lyapunov equation*. The control law follows to

$$\boldsymbol{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t). \tag{6.81}$$

Example 6.8

The following state-space system is considered:

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u.$$

It is apparent that the second state x_2 is unstable ($\lambda_2 = 1$) and uncontrollable since $x_2(t)$ is neither directly nor indirectly affected by $u(t)$. Consequently, the system is not stabilizable. Choosing the weighting matrices to

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 1,$$

the Riccati equation (6.78) is

$$-\dot{\boldsymbol{P}} = \boldsymbol{P}\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}^T\boldsymbol{P} - \boldsymbol{P}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{P} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Figures 6.9 and 6.10 show the components $P_{11}(t)$, $P_{12}(t) = P_{21}(t)$ and $P_{22}(t)$, if the Riccati equation is solved backward in time starting from $\boldsymbol{P}(3) = \boldsymbol{0}$. While $P_{11}(t)$ and $P_{12}(t)$ quickly converge toward their steady-state equilibrium, the component $P_{22}(t)$ is obviously unbounded.

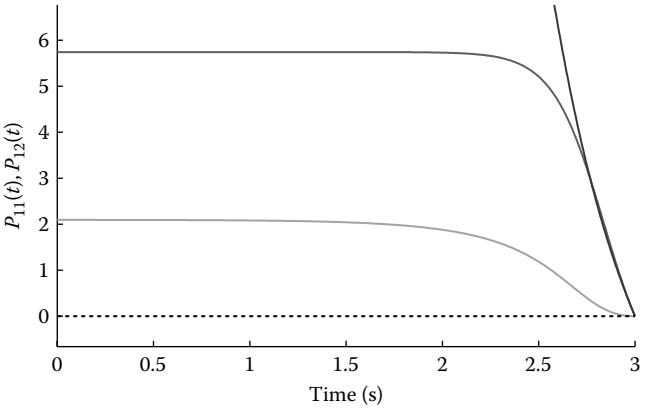


FIGURE 6.9 Example 6.8: $P_{11}(t)$ and $P_{12}(t)$.

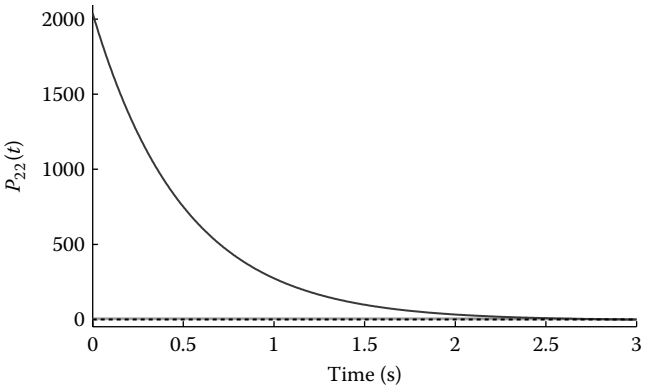


FIGURE 6.10 Example 6.8: $P_{22}(t)$.

6.3.3 *Q–R* Selection

This section will provide different criteria for the proper choice of the weighting matrices.

6.3.3.1 Basics

The elements of the state vector and the control inputs can vary significantly in size. A basic rule of thumb is to choose $\boldsymbol{Q}$ and $\boldsymbol{R}$ as diagonal matrices with

$$q_{ii} = \frac{1}{(x_{i,max})^2} \quad \text{and} \quad r_{ii} = \frac{1}{(u_{i,max})^2},$$

where $x_{i,max}$ and $u_{i,max}$ are the maximum desired amplitudes of the individual states and controls, respectively.

6.3.3.2 State Weighting versus Control Weighting

Strong state weighting. If the elements of $\boldsymbol{Q}$ are large compared to those of $\boldsymbol{R}$, there is a strong emphasis on the states in the cost function. The resulting controller will consequently aim at bringing the states to zero as fast as possible. As a result, the control effort will increase and the bandwidth of the closed loop will grow.

Strong control weighting. If the elements of $\boldsymbol{Q}$ are small compared to those of $\boldsymbol{R}$, there is a strong emphasis on the controls in the cost function. Consequently, the controller will try to reduce the control effort as much as possible. For stable systems, this will result in a control law $u \equiv 0$ and the poles of the closed loop will approach those of the open loop. For systems with unstable poles, these poles will be reflected on the imaginary axis resulting in a stable closed-loop system.

6.3.3.3 Stability

Stability of the closed loop mainly depends on the proper choice of the weighting matrix $\boldsymbol{Q}$.

Theorem 6.7 (Detectability)

*The matrices [**A**, **Q**] are called detectable, if for **A** every eigenvalue λ_i with positive real part the eigenvector $\boldsymbol{u}_i$ fulfills*

$$\boldsymbol{u}_i^T \boldsymbol{Q} \boldsymbol{u}_i \neq \boldsymbol{0}. \tag{6.82}$$

In the sense of optimal control, detectability means that unstable eigenmodes of the systems show up in the cost function.

The following conditions have to be fulfilled for the closed loop to be stable:

- Uncontrollable modes of the system $[\boldsymbol{A},\boldsymbol{B}]$ are stable, that is, $[\boldsymbol{A},\boldsymbol{B}]$ is *stabilizable* (cf. Theorem 6.6).
- Unstable modes of the system are visible in the cost function, that is, $[\boldsymbol{A},\boldsymbol{Q}]$ is detectable.

Example 6.9

The following system is considered:

$$\dot{x} = x + u.$$

If the cost function is chosen as

$$V_*(\boldsymbol{x}(t),\ \boldsymbol{u}_*(t)) = \int_0^\infty u^2 dt.$$

It is apparent that the optimal control law is $u \equiv 0$. The closed loop would then be unstable although the system is fully state controllable because the unstable mode does not show up in the cost function.

6.3.3.3.1 Assigning a Predefined Degree of Stability

Due to robustness requirements, it is often desirable to ensure a certain degree of stability, that is, the real parts of the closed-loop poles must be smaller than some predefined constant α.

The open-loop system (6.75) is

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} \qquad \boldsymbol{x}(t=0) = \boldsymbol{x}_0\,.$$

For the cost function, new state and control variables $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{u}}$ are introduced:

$$\hat{\boldsymbol{x}} = e^{\alpha t}\boldsymbol{x} \quad \text{and} \quad \hat{\boldsymbol{u}} = e^{\alpha t}\boldsymbol{u}. \tag{6.83}$$

Differentiating and inserting $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$ gives

$$\dot{\hat{\boldsymbol{x}}} = \frac{d}{dt}(e^{\alpha t}\boldsymbol{x}) = (\boldsymbol{A} + \alpha\boldsymbol{I})\hat{\boldsymbol{x}} + \boldsymbol{B}\hat{\boldsymbol{u}}. \tag{6.84}$$

The cost function (6.76) is now expressed in terms of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{u}}$:

$$V(\boldsymbol{x}(0),\ \boldsymbol{u}(\cdot)) = \int_0^\infty \frac{1}{2}(\hat{\boldsymbol{x}}^T\boldsymbol{Q}\hat{\boldsymbol{x}} + \hat{\boldsymbol{u}}^T\boldsymbol{R}\hat{\boldsymbol{u}})\,dt. \tag{6.85}$$

If the closed loop should have a degree of stability α, a controller for modified system (6.84) based on the cost function (6.85) must be stable. From Theorems 6.6 and 6.7, this entails

1. $[\boldsymbol{A} + \alpha\boldsymbol{I}, \boldsymbol{B}]$ is stabilizable.
2. $[\boldsymbol{A} + \alpha\boldsymbol{I}, \boldsymbol{Q}]$ is detectable.

If $\hat{\boldsymbol{P}}$ is the solution of the Lyapunov equation associated with 6.84 and 6.85 according to

$$\hat{\boldsymbol{P}}(\boldsymbol{A}+\alpha\boldsymbol{I})+(\boldsymbol{A}+\alpha\boldsymbol{I})^T\hat{\boldsymbol{P}}-\hat{\boldsymbol{P}}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\hat{\boldsymbol{P}}+\boldsymbol{Q}=\boldsymbol{0}, \tag{6.86}$$

then the control law is

$$\boldsymbol{u}(t)=e^{-\alpha t}\hat{\boldsymbol{u}}(t)=-e^{-\alpha t}\boldsymbol{R}^{-1}\boldsymbol{B}^T\hat{\boldsymbol{P}}e^{\alpha t}\boldsymbol{x}(t)=-\boldsymbol{R}^{-1}\boldsymbol{B}^T\hat{\boldsymbol{P}}\boldsymbol{x}(t). \tag{6.87}$$

6.3.3.4 Modal Weighting

In many applications, it is useful if certain undesirable states of the system could be particularly weighted in the cost function. For example, these could be oscillations or unstable eigenmodes that could destabilize the control loop. For this purpose, the matrix $\boldsymbol{Q}$ can be designed such that these modes have a predefined effect on the cost function.

Let $\boldsymbol{s}_1 \ldots \boldsymbol{s}_n$ be the unit eigenvectors of the open-loop system. If there is a complex conjugate eigenvalue couple, then the corresponding eigenvector has to be separated into real and imaginary parts.

Next, an auxiliary matrix $\hat{\boldsymbol{Q}}$ is designed such that for every vector $\boldsymbol{s}_k$, $k = 1 \ldots n$ the following relation must hold:

$$\hat{\boldsymbol{Q}}\boldsymbol{s}_k=\boldsymbol{s}_k m_k, \quad k=1\ldots n. \tag{6.88}$$

Here, m_k is a positive real weighting factor. If all m_k are collected in a diagonal matrix $\boldsymbol{M}$ and all $\boldsymbol{s}_k$ are concatenated to a matrix $\boldsymbol{S}$, then 6.88 can be collected for all $k = 1 \ldots n$:

$$\hat{\boldsymbol{Q}}\boldsymbol{S}=\boldsymbol{S}\boldsymbol{M}. \tag{6.89}$$

The auxiliary matrix $\hat{\boldsymbol{Q}}$ is obtained from

$$\hat{\boldsymbol{Q}}=\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}^{-1} \tag{6.90}$$

and the weighting matrix $\boldsymbol{Q}$ is obtained from

$$\boldsymbol{Q}=\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{Q}}. \tag{6.91}$$

The last step ensures that $\boldsymbol{Q}$ is symmetric. If $\boldsymbol{s}_k$ is inserted into the cost function, one obtains

$$\boldsymbol{s}_k^T\boldsymbol{Q}\boldsymbol{s}_k=\boldsymbol{s}_k^T\hat{\boldsymbol{Q}}^T\hat{\boldsymbol{Q}}\boldsymbol{s}_k=m_k^2 \quad \text{with } \|\boldsymbol{s}_k\|=1. \tag{6.92}$$

It is apparent that the kth eigenmode contributes to the cost function proportionally to m_k^2.

Modal weighting can be successfully applied to design controllers that have to fulfill a very specific control task. Typically, m_k is chosen small for nonrelevant modes whereas large values of m_k are assigned to the undesirable modes.

6.3.3.5 Weighting of State Derivatives

Often state derivatives (e.g., accelerations) have to be considered in the cost function. This section shortly describes how this can be accomplished.

Consider the following cost function:

$$V=\int_0^\infty\frac{1}{2}(\dot{\boldsymbol{x}}^T\boldsymbol{Q}\dot{\boldsymbol{x}}+\boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u})\,dt. \tag{6.93}$$

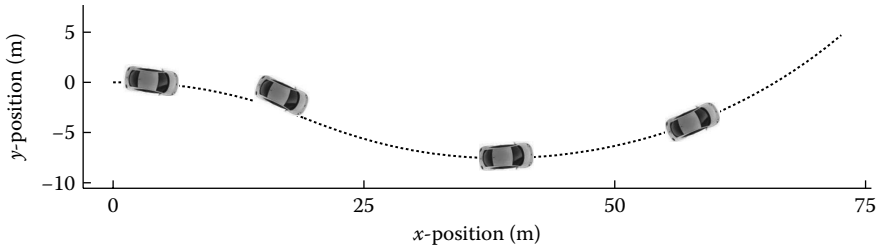


FIGURE 6.11 C/G trajectory with LQR control.

Inserting $\dot{\boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{Bu}$ leads to

$$V = \int_0^\infty \frac{1}{2}(\boldsymbol{x}^T \bar{\boldsymbol{Q}}\boldsymbol{x} + 2\boldsymbol{x}^T \boldsymbol{N}\boldsymbol{u} + \boldsymbol{u}^T \bar{\boldsymbol{R}}\boldsymbol{u})dt \tag{6.94}$$

with $\bar{\boldsymbol{Q}} = \boldsymbol{A}^T\boldsymbol{QA}$, $\bar{\boldsymbol{R}} = \boldsymbol{B}^T\boldsymbol{QB} + \boldsymbol{R}$ and $\boldsymbol{N} = \boldsymbol{A}^T\boldsymbol{QB}$. The control law then takes the form

$$\boldsymbol{u}(t) = -\bar{\boldsymbol{R}}^{-1}(\boldsymbol{N}^T + \boldsymbol{B}^T\bar{\boldsymbol{P}})\boldsymbol{x}(t) \tag{6.95}$$

where $\bar{\boldsymbol{P}}$ is the solution of a modified Lyapunov equation:

$$\bar{\boldsymbol{P}}(\boldsymbol{A} - \boldsymbol{B}\bar{\boldsymbol{P}}^{-1}\boldsymbol{N}^T) + (\boldsymbol{A} - \boldsymbol{B}\bar{\boldsymbol{P}}\boldsymbol{N}^T)^T\bar{\boldsymbol{P}} - \bar{\boldsymbol{P}}\boldsymbol{B}\bar{\boldsymbol{R}}^{-1}\boldsymbol{B}^T\bar{\boldsymbol{P}} + (\boldsymbol{Q} - \boldsymbol{N}\bar{\boldsymbol{R}}^{-1}\boldsymbol{N}^T)\boldsymbol{0}. \tag{6.96}$$

Example 6.10

Again, the single-track model from Example 6.1 is considered.

In contrast to Example 6.5, the steering angle is now considered as an additional control input. The new control input can be considered as an additional steering wheel angle $\Delta\delta$ that is superimposed to the driver's input.

The selection of the weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ is based on the simple criteria from Section 6.3.3.1.

$\boldsymbol{Q}$ is chosen as the identity matrix, the control weights are chosen as

$$\boldsymbol{R} = \begin{bmatrix} 5000 & 0 \\ 0 & 10^{-7} \end{bmatrix}. \tag{6.97}$$

Figure 6.11 shows the trajectory of the car under the respective control action. Figure 6.12 illustrates the activity of the superimposed control $\Delta\delta$ and Figure 6.13 shows a time history of the second control input (M). A comparison of Figures 6.5 and 6.11 reveals that the stabilization by the controller is now somewhat slower. However, the bandwidth LQR controller could be readily improved by increasing the entries of the $\boldsymbol{Q}$ matrix.

6.4 IDENTIFICATION—EXPERIMENTAL MODELING

6.4.1 Introduction

System identification is the art of extracting both structural and quantitative knowledge about a dynamic system from phenomenological data only. For a given model, structure methods exist to estimate a set of optimal parameters. Nevertheless, the model structure itself has to be defined a priori by the user, and optimal structure can only be shown a posteriori using validation methods. Therefore, system identification is always an iterative process.

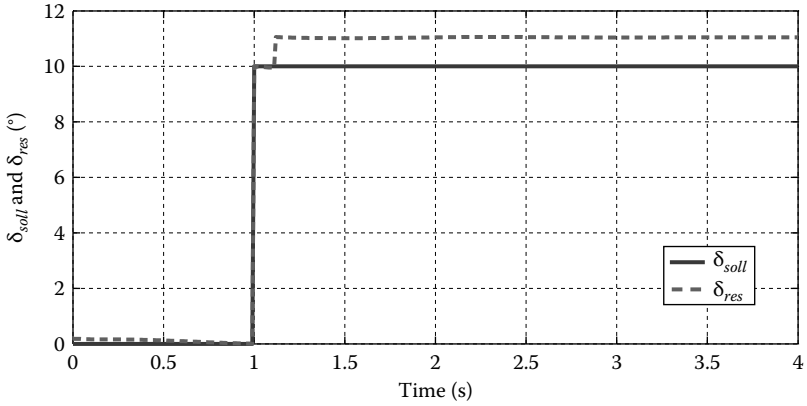


FIGURE 6.12 Time history of δ_{soll} and resultant δ_{res} ($\delta_{res} = \delta_{soll} + \Delta\delta$).

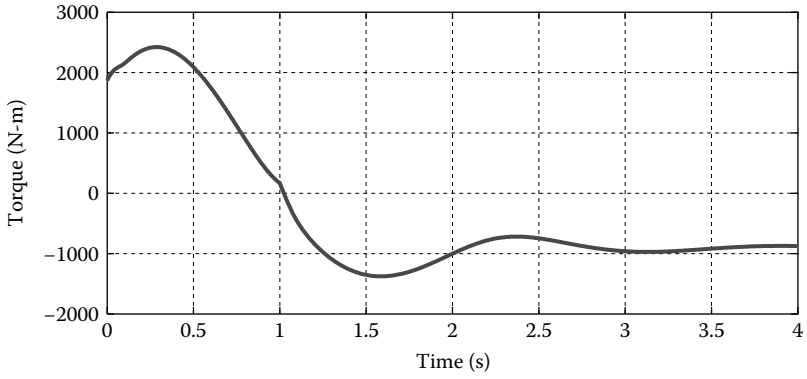


FIGURE 6.13 Time history of control input two (torque).

The components of an identification problem are

- Measured input and output signals
- General structural assumptions (e.g., linearity, certain polynomial orders, etc.)
- A number of system parameters that parameterize the model structure and uniquely define the system's behavior

The procedure of an identification is closely intertwined with the important part of model validation (see Figure 6.14). To this end, the data are divided in an identification and a validation part, respectively. Assuming a certain model structure, the model parameters can be estimated utilizing the identification

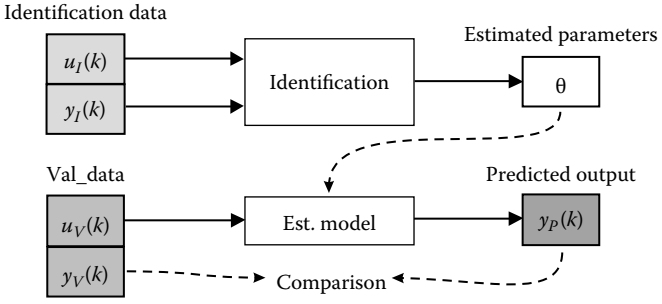


FIGURE 6.14 Identification and cross-validation of an experimental model.

data and a suitable estimation method. The quality of this model is consequently tested on the validation data, where the ability to correctly predict the measured output is an important criterion.

The remainder of this section is structured as follows: first, some basic notions on discrete-time signals and systems are introduced. This is followed by an introduction to probability and stochastics, as the treatment of random measurement noise requires some of these concepts. The core part is formed by an introduction to nonparametric and parametric estimation methods, where mainly linear systems are addressed. This is followed by short step-by-step instructions how to accomplish a successful model identification.

Because of the close affinity to identification procedures, the Kalman filter and its subsequent developments as well as alternative methods for state estimation are also included in this section.

6.4.2 Fundamentals

6.4.2.1 Reality, System, and Model

The notion of a *system* is used to describe a set of interacting elements inside a system boundary, which may exchange energy or information with the environment by means of input and output signals (Figure 6.14). A large or complex system can be divided into subsystems, where the inner interactions of the original system become visible. On the other hand, a set of subsystems can be combined in one complete system and the inner interactions are no longer explicit (Figure 6.15).

In order to represent a well-defined domain of a technical application by a system (this mapping will be called *model*), the system boundaries have to be chosen carefully. This is especially important, since the input and output signals necessary for identification are also defined with the choice of the boundary. The ability and performance of an experimental model critically rely on the choice of the system boundaries and hence the input and output signals [14].

6.4.3 Model Validation

An experimental model is only as good as its validation. The task of a model validation is a realistic performance test in order to guarantee an acceptable amount of reliability and accuracy. The method of cross-validation is the most reliable way to validate a model: measured data are split into a larger identification part and a smaller validation part (Figure 6.14). Since the model is validated with data that have not been used during identification, this test constitutes a simulation of the model's real-world application. It tests the model's ability to reproduce not only the identification data but also more general data (this ability is called *generalization* [7,13,14]).

Model validation is performed in a recursive manner (Figure 6.16). Since assumptions on model structure have to be made prior to the identification procedure, each identification has to be followed by a validation. The results of the validation indicate which alterations in the model are necessary (structure, simplifications, parameters). The iteration is stopped when the trade-off between model

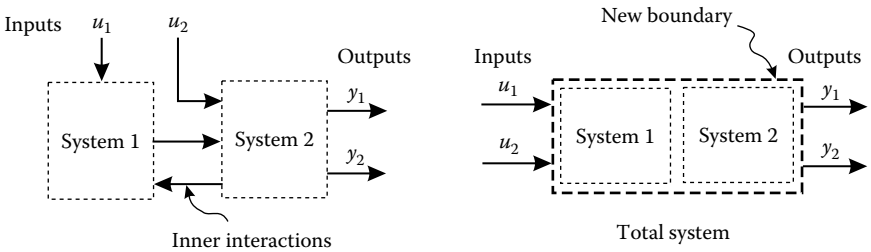


FIGURE 6.15 A complete system can be divided into a number of subsystems. The inner interactions are not explicitly visible in the original system.

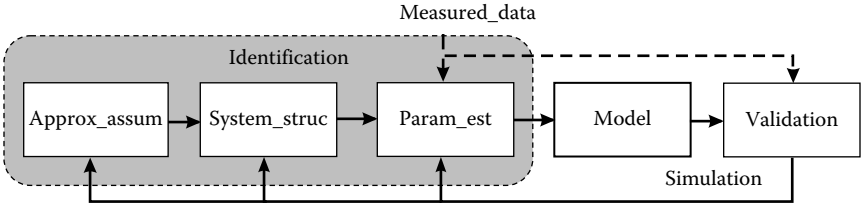


FIGURE 6.16 Identification and validation are an iterative process.

accuracy and complexity comes to an optimum. Depending on the purpose of an experimental model, accuracy may be less important than stability, simplicity, and generalization.

6.4.4 Basics of Discrete-Time Signals and Systems

Discrete-time signals are built from continuous-time signals using a procedure called "sampling" [18,19]. The time axis is divided into equidistant sections (sampling time T_s), and signal values are only stored ("sampled") at the intersections (see Figure 6.17a). Each time instant of the sampled signal $x(kT_s)$ is denoted by a dimensionless counter k:

$$x(t)\big|t = kT_s = x(kT_s) = x(k), \quad k \in Z. \tag{6.98}$$

The values of the original signal between the sampling instants can only be reconstructed up to the Nyquist frequency ω_ν [18,19], where

$$\omega_\nu = \frac{1}{2}\omega_s = \frac{1}{2}2\pi f_s = \frac{\pi}{T_s} \tag{6.99}$$

holds. An analog low-pass filter has to be employed prior to sampling in order to avoid aliasing effects. Discrete-time system descriptions consist of difference equations as opposed to differential equations for continuous-time systems. It is important to note that the sampling time T_s is inherently stored in the system parameters and acts as a scaling parameter on the time axis.

Example 6.11

The continuous-time model of a damped linear oscillator with external force F(t) is given by the differential equation:

$$\ddot{x}(t) + 1.2\dot{x}(t) + 9x(t) = F(t).$$

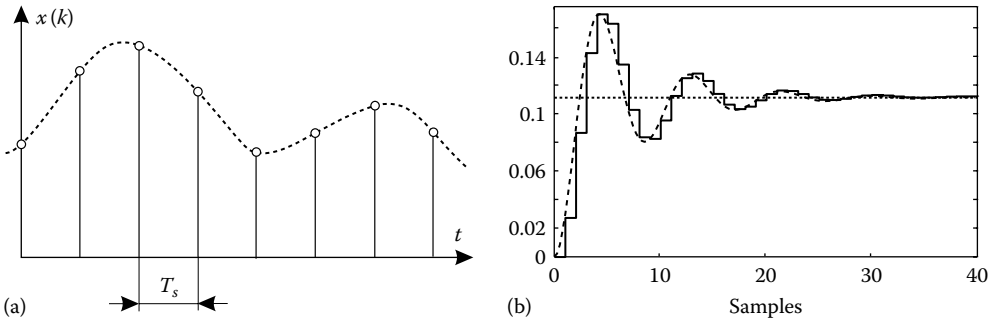


FIGURE 6.17 (a) Sampling of a continuous-time signal with a sample time of T_s. (b) Step response for continuous-time and discrete-time model of a damped oscillator.

Discretization with a sampling time of $T_s = 0.25$ s yields a difference equation:

$$x(k) - 1.2772x(k-1) + 0.7408x(k-2) = 0.0271F(k-1) + 0.0244F(k-2).$$

In Figure 6.17b, the step responses of both systems are plotted. Note that the response of the sampled system is identical to the continuous one only at the sampling instants; during the sampling interval, the output is kept constant which indicates the loss of information.

6.4.4.1 Pulse Transfer Function

A discrete-time linear dynamic SISO system is defined through a difference equation:

$$d_n y(k-n) + \cdots + d_1 y(k-1) + d_0 y(k) = c_m u(k-m) + \cdots + c_1 u(k-1) + c_0 u(k). \tag{6.100}$$

Assuming vanishing initial conditions ($u(k) = y(k) = 0, \forall k \leq 0$) and using the z-transform [18,19], the difference equation (6.100) can be transformed into

$$d_n z - {}^n Y(z) + \cdots + d_1 z^{-1} Y(z) + d_0 Y(z) = c_m z^{-m} U(z) + \cdots + c_1 z^{-1} U(z) + c_0 U(z). \tag{6.101}$$

The discrete-time pulse transfer function in the complex variable z is defined as

$$G(z) = \frac{Y(z)}{U(z)} = \frac{c_m z^{-m} + \cdots + c_1 z^{-1} + c_0}{d_m z^{-n} + \cdots + d_1 z^{-1} + d_0}. \tag{6.102}$$

The parameters c_i and d_j define the dynamic properties of the system [18,19].

6.4.5 Basics of Probability and Stochastics

6.4.5.1 Probability of an Event

An event A is a well-defined subset of all possible outcomes a of a random process.

Example 6.12

A random process has possible outcomes of $0 \leq a < 360$. One can define the events $A_1 = \{120 \leq a < 180\}$ and $A_2 = \{300 \leq a < 360\}$.

If a random process is sampled N times and the absolute frequency of the event A is given by $N[A]$, the relative frequency of the event A is defined as $n[A] = N[A]/N$. The limit

$$P[A] = \lim_{N \to \infty} n[A] = \lim_{N \to \infty} \frac{N[A]}{N} \tag{6.103}$$

empirically defines the probability of the event A.

6.4.5.2 Definition of a Random Variable

A random variable X is a real-valued function that maps all samples of a random process on the real axis or parts thereof [3]. It must not be multi-valued and can therefore be seen as a universal description of any random process.

6.4.5.2.1 Probability Density Function

Due to the earlier defined mapping, the probability of an event $P[A]$ with A given by $A = \{a \leq x < b\}$ is

$$P[A] = P[a \leq x < b] \tag{6.104}$$

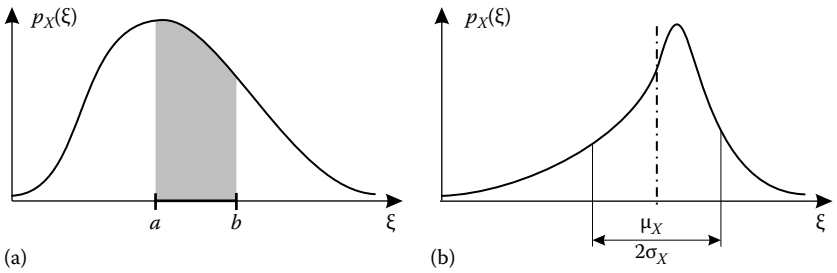


FIGURE 6.18 Continuous probability density function (PDF) $p_x(\xi)$. (a) Probability that the value of the random variable falls in an interval $a \le X \le b$. (b) Mean and standard deviation constitute the weak description of a random variable.

and can be expressed using the *probability density function* (PDF) $p_X(\xi)$ (see Figure 6.18a):

$$P[A] = P[a < X \le b] = \int_a^b p_X(\xi) d\xi. \tag{6.105}$$

Properties of the PDF $p_X(\xi)$:

$$0 \le p_X(\xi),$$

$$p_X(-\infty) = p_X(\infty) = 0,$$

$$\int_{-\infty}^{\infty} p_X(\xi)\, d\xi = 1. \tag{6.106}$$

The PDF is the strict sense description of a random variable.

6.4.5.2.2 Expectation and Moments—Mean and Variance

The *expectation* or (linear) *mean* μ_X of a random variable X is defined as the first-order moment of the WDF $p_X(\xi)$ [3],

$$\hat{\mu}_x = E\{X\} = \bar{X} = \int_{-\infty}^{\infty} \xi p_X(\xi) d\xi. \tag{6.107}$$

In Equation 6.107, the *expectation operator* $E\{\}$ assigns the deterministic expectation μ_X to the random variable X. The empirical mean $\hat{u}_x$ of a discrete-time signal $x(k)$ with N samples is given by

$$\hat{\mu}_x = \frac{1}{N} \sum_{k=1}^{N} x(k).$$

Note that the empirical mean $\hat{\mu}_x$ depends on the sample values $x(k)$ and therefore is in itself a random variable while μ_X is the true (deterministic) mean of the underlying random process. The *variance* σ_X^2 is defined as the mean quadratic deviation of the random variable X from its mean value μ_X (second centered moment),

$$\sigma_X^2 = E\{X - \mu_X)^2\} = \int_{-\infty}^{\infty} (\xi - \mu_X)^2 p_X(\xi) d\xi.$$

The empirical variance $\hat{\sigma}_x^2$ of a discrete-time signal $x(k)$ with N samples is given by

$$\hat{\sigma}_x^2 = \frac{1}{N}\sum_{k=1}^{N}(x-\hat{\mu}_x)^2(k).$$

The square root of the variance is called *standard deviation* σ_X (see Figure 6.18b). Mean and variance constitute the weak description of a random variable. Using the expectation operator, the weak description can always be computed from the PDF but not vice versa. Note that all empirical estimates are random variables in themselves; this implies that they have a mean and variance depending on the number of samples used for computation N as well as the reliability of the sampled data (measurement noise)!

6.4.5.2.3 Correlation and Covariance

If two random variables X_1, X_2 with a joint probability density function (JPDF) $p_{X_1X_2}(\xi_1, \xi_2)$ are considered simultaneously, the means μ_{X_1}, μ_{X_2} and variances $\sigma_{X_1}^2$, σ_{X2}^2 contain no information about their statistical dependency. Therefore, the *correlation* $p_{X_1X_2}^2$ and the *covariance* $\sigma_{X_1X_2}^2$ are defined [3,6]:

$$p_{X_1X_2}^2 = E\{X_1X_2\}, \quad \sigma_{X_1X_2}^2 = E\{(X_1-\mu_{X_1})(X_2-\mu_{X_2})\}. \tag{6.108}$$

Two random variables are *uncorrelated* if

$$\sigma_{X_1X_2}^2 = 0 \quad \text{or} \quad p_{X_1X_2}^2 = \mu_{X_1}, \mu_{X_2} \quad \text{or} \quad E\{X_1X_2\} = E\{X_1\}E\{X_2\}.$$

In this case, $\sigma_{X_1+X_2}^2 = \sigma_{X_1}^2 + \sigma_{X2}^2$ holds. For practical applications, the empirical correlation $\hat{p}_{xy}^2$ and covariance $\hat{\sigma}_{xy}^2$ between two discrete-time signals $x(k)$ and $y(k)$ are a measure for statistical dependence:

$$\hat{\rho}_{xy}^2 = E\{x(k)y(k)\} = \frac{1}{N}\sum_{k=1}^{N}x(k)y(k),$$

$$\hat{\sigma}_{xy}^2 = E\{[x(k)-\hat{\mu}_x][y(k)-\hat{\mu}_y]\} = \frac{1}{N}\sum_{k=1}^{N}[x(k)-\hat{\mu}_x][y(k)-\hat{\mu}_y].$$

In order to have a normalized measure for correlation, the *correlation coefficient* $\hat{r}_{xy}$ is defined as

$$\hat{r}_{xy} = \frac{\hat{\sigma}_{xy}^2}{\sqrt{\hat{\sigma}_x^2\hat{\sigma}_y^2}} = \frac{\hat{\sigma}_{xy}^2}{\hat{\sigma}_x\hat{\sigma}_y}, \quad |\hat{r}_{xy}| \le 1.$$

The strict sense description of two random variables X_1X_2 is given by the JPDF; the wide sense description is given by the means μ_{X_1}, μ_{X_2}, the variances $\sigma_{X_1}^2, \sigma_{X_2}^2$, and the covariance $\sigma_{X_1+X_2}^2$ (instead of $\sigma_{X_1+X_2}^2$, $p_{X_1X_2}^2$ can also be used).

6.4.5.3 Covariance Matrix

The covariance matrix $\hat{\boldsymbol{C}}$ is a quadratic scheme to clearly describe the statistical dependencies between n different signals [6,13]. The individual signals $x_i(k)$ are combined into a measurement

matrix $\boldsymbol{X} = [x_1 x_2 \ldots x_n]$ (each column constitutes a signal; each line corresponds to a sampling instant k). The *covariance matrix* is defined as

$$\hat{\boldsymbol{C}}(\boldsymbol{x}) = E\{[\boldsymbol{x}(k) - \hat{\boldsymbol{\mu}}_x][\boldsymbol{x}(k) - \hat{\boldsymbol{\mu}}_x]^T\} = \begin{bmatrix} \hat{\sigma}^2_{x_1 x_1} & \hat{\sigma}^2_{x_1 x_2} & \cdots & \hat{\sigma}^2_{x_1 x_n} \\ \hat{\sigma}^2_{x_2 x_1} & \hat{\sigma}^2_{x_2 x_2} & \cdots & \hat{\sigma}^2_{x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}^2_{x_n x_1} & \hat{\sigma}^2_{x_n x_2} & \cdots & \hat{\sigma}^2_{x_n x_n} \end{bmatrix}$$

The vector $\hat{\boldsymbol{\mu}}_x = [\hat{\mu}_x, \hat{\mu}_2, \ldots, \hat{\mu}_n]$ contains the means of the individual signals. The covariance matrix is a quadratic, symmetric, positive (semi-)definite $n \times n$ matrix, whose elements consist of covariances and variances (in the diagonal), respectively:

$$\hat{\sigma}^2_{x_i x_j} = E\{[x_i(k) - \hat{\mu}_{x_i}][x_j(k) - \hat{\mu}_{x_j}]\}, \quad \hat{\sigma}^2_{xixi} = E\{[x_i(k) - \hat{\mu}_{x_i}]^2\}.$$

The covariance matrix $\hat{\boldsymbol{C}}$ is positive definite if and only if all signals $x_i(k)$ are linear-independent random variables.

6.4.5.4 Auto- and Cross-Correlation Functions

The correlation between two signals $x(k)$ and $y(k)$ provides scalar information about statistical dependencies. However, depending on a relative time shift τ between the two signals and the signal characteristic, the correlation may change considerably with τ. Inserting the time shift into Equation 6.108 yields

$$R_{xy}(\tau) = E\{x(k)y(k-\tau)\}, \tag{6.109}$$

the cross-correlation function (CCF) between $x(k)$ and $y(k)$. The CCF is dependent only on τ if the signals $x(k)$ and $y(k)$ are stationary (meaning that their statistical properties are invariant to a time shift). Important properties of the CCF are [2,3,19]

$$R_{xy}(\tau) = R_{yx}(-\tau), \tag{6.110}$$

$$R_{xy}(0) = E\{x(k)y(k)\} = \rho^2_{xy}. \tag{6.111}$$

The autocorrelation function (ACF) is defined as

$$R_{xx}(\tau) = E\{x(k)x(k-\tau)\}, \tag{6.112}$$

with the properties

$$R_{xx}(\tau) = R_{xx}(-\tau), \tag{6.113}$$

$$R_{xx}(0) = E\{x(k)x(k)\} = \sigma^2_x. \tag{6.114}$$

The empirical ACF and CCF for signals $x(k)$ and $y(k)$ with N samples are given by

$$\text{ACF} \qquad \hat{R}_{xx}(l) = F_s \sum_{k=1}^{N} x(k)x(k-l),$$

$$\text{CCF} \qquad \hat{R}_{xy}(l) = F_s \sum_{k=1}^{N} x(k)y(k-l),$$

where $F_s = 1/(N - |l|)$ is a scaling factor for an unbiased estimate (see Section 6.5.3.2).

6.5 IDENTIFICATION METHODS

6.5.1 Input Signals and Persistent Excitation

The choice of a proper input signal is important since the unknown system dynamics have to be excited in order to show up in the measured outputs. A Dirac-delta impulse $u(t)=\delta(t-t_0)$ will certainly excite the system's dynamics; however, energy is fed into the system only at a single time instant t_0. On the other hand, a harmonic signal $u(t) = \sin(\omega_0 t)$ will permanently excite the system but only at a single frequency ω_0. It would therefore be desirable to have a signal which combines both advantages. This signal is a purely random sequence where constant power is contained over the frequency range. It is called *white noise* [3,13].

A necessary condition for a system of order $n \leq m$ to be identifiable is that the input is *persistently exciting* (PE) of order $2m$. The Dirac-delta impulse is PE of order 0, a signal consisting of a sum of m sine functions is PE of order $2m$, and white noise is PE of order ∞ [13].

Example 6.13

The ACF of a white random signal w(k) with variance σ_w^2 is given by $R_{\omega\omega}(\tau)$. Here, δ_0 denotes the Kronecker-delta function, which is 1 at $\tau = 0$ and zero elsewhere. Two white noise sequences are given by w(k) and v(k) = w(k – d), that is, v(k) is a version of w(k) shifted by d samples. Both have the same ACF given in the earlier example and they have a CCF of $R_{wv}(\tau) = \sigma_w^2\delta_0(\tau - d)$.

6.5.2 Nonparametric Modeling

Nonparametric models utilize characteristic functions in the time and frequency domains to describe a dynamic system. They are compact and robust, but yield mostly qualitative results and are not suitable for direct simulation. One of the main advantages is that only minimal assumptions on the model order or structure are necessary: linearity, uncorrelated measurement noise (no applicability in closed loop), and PE input.

6.5.2.1 Step Response

According to the definition of the step response, a stepwise input $u(k) = \sigma(k)$ is fed into the system. The output can be seen as a nonparametric model of the process. Stationary gain, dead time, and characteristic time constants can be extracted from the output. Furthermore, for MIMO systems a quick overview over couplings between inputs and outputs is possible [13,14]. Instead of the step, an impulse can be used as an alternate excitation.

6.5.2.2 Correlation Analysis

Correlation analysis is a method for estimating the impulse or step response for nearly arbitrary input signals. Furthermore, statistical reliability is guaranteed even in the presence of measurement noise. The method is based on the Wiener–Hopf equation:

$$R_{uy}(k) = \sum_{h=0}^{\infty} g(h) R_{uu}(h+k), \tag{6.115}$$

where

R_{uy} is the CCF between input u and output y
R_{uu} is the ACF of the input
$g(h)$ is the impulse response of the system

If the input signal is white noise with variance σ_u^2, Equation 6.115 can be solved for $g(k)$:

$$g(k) = \frac{1}{\sigma_u^2} R_{uy}(k). \tag{6.116}$$

Since Equation 6.116 is only true for a white input sequence, an additional procedure is necessary for arbitrary signals. A so-called prewhitening filter is estimated such that the spectrum of the real input is equal to white noise filtered through the inverse prewhitening filter. Using the filtered versions of the measured signals, an unbiased estimate of $g(k)$ is possible [2,13,14,19].

6.5.2.3 Spectral Analysis

The spectrum $\Phi_u(i\omega)$ of a signal $u(k)$ describes the average power content depending on the frequency ω. The complex-valued estimate $\hat{\Phi}_u(i\omega)$ (periodogram) is typically computed utilizing an averaging method in order to achieve a consistent but biased (methods by Welch or Blackman–Tukey).

The nonparametric estimate of the transfer function $\hat{G}(i\omega)$ is then computed by

$$\hat{G}(i\omega) = \frac{\hat{\Phi}_{yu}(i\omega)}{\hat{\Phi}_u(i\omega)}, \tag{6.117}$$

where $\hat{\Phi}_{yu}(i\omega)$ denotes the cross-spectral density estimate between $u(k)$ and $y(k)$. Depending on the implementation, some additional numerical parameters have to be chosen. Nevertheless, there is always a dependency on a smoothing parameter. Strong smoothing will give a biased but consistent estimate, while no smoothing will yield an unbiased but very noisy estimate [2,14,19].

6.5.3 Parametric Modeling

In this section, the basic notions of linear parameter estimation will be introduced. The models will have mainly the structure of a discrete transfer function, but any other structure, which is linear in the parameters, can be treated in the same way.

6.5.3.1 Problem Formulation

In parametric models, the presence of measurement noise (see Figure 6.19) has to be explicitly addressed by the estimation procedure. The measured output is given by

$$y(k) = y_0(k) + n(k).$$

A parametric model in the formulation of a discrete-time linear difference equation is defined by

$$d_n y(k-n) + \cdots + d_1 y(k-1) + d_0 y(k) = c_m u(k-m) + \cdots + c_1 u(k-1) + c_0 u(k).$$

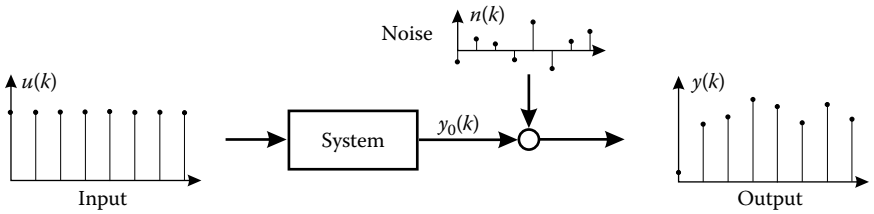


FIGURE 6.19 The measured output $y(k)$ is composed of the deterministic output $y_0(k)$ plus the measurement noise $n(k)$.

Rearranging for the current output $y(k)$ given by past outputs $y(k-i)$ and inputs $u(k-j)$ yields

$$y(k) = a_n y(k-n) + \cdots + a_1 y(k-1) + \beta_m u(k-m) + \cdots + \beta_1 u(k-1) + \beta_0 u(k). \tag{6.118}$$

Changing into matrix-vector representation gives

$$y(k) = [y(k-n)\ldots y(k-1)u(k-m)\ldots u(k-1)u(k)] \begin{bmatrix} \alpha_n \\ \vdots \\ \alpha_1 \\ \beta_m \\ \vdots \\ \beta_1 \\ \beta_0 \end{bmatrix}, \tag{6.119}$$

which can be augmented for all time instances k:

$$\begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(k+N) \end{bmatrix} = \begin{bmatrix} y(k-n) & \cdots y(k-1) & u(k-m) & \cdots u(k-1) & u(k) \\ y(k-n+1) & \cdots y(k) & u(k-m)+1 & \cdots u(k) & u(k+1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y(k-n+N) & \cdots y(k-1+N) & u(k-m+N) & \cdots u(k-1+N) & u(k+N) \end{bmatrix} \begin{bmatrix} \alpha_n \\ \vdots \\ \alpha_1 \\ \beta_m \\ \vdots \\ \beta_1 \\ \beta_0 \end{bmatrix}. \tag{6.120}$$

Writing 6.120 in symbolic form yields

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}. \tag{6.121}$$

where

- $\boldsymbol{y}$ denotes the output vector
- $\boldsymbol{X}$ is the regressor matrix
- $\boldsymbol{\theta}$ is the parameter vector

It should be observed that the system order (n and m, respectively) has to be chosen prior to constructing the regressor matrix. This means that the optimal system order has to be found iteratively, since each solution to 6.121 is only valid for a dedicated combination of n and m, respectively.

Remark:

A dead time of d samples can be easily incorporated by replacing the arguments (k) of y in the regressor matrix (6.120) by ($k-d$).

Equation 6.121 defines a system of linear equations; solving for $\boldsymbol{\theta}$ would give the desired solution to the parameter estimation problem. A formal requirement for (6.121) to have a solution is a quadratic non-singular regressor matrix $\boldsymbol{X}$. However, the regressor matrix in general has many more lines (number of samples) than columns (number of parameters). Obviously, some alternate method has to be employed in order to solve the overdetermined system of equations. This leads to the following section.

6.5.3.2 Least-Squares Algorithm

Considering the presence of measurement noise, the output vector $\boldsymbol{y}$ can be expressed using the regression matrix $\boldsymbol{X}$, the parameter vector $\boldsymbol{\theta}$, and the equation error $\boldsymbol{e}$:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{e}. \tag{6.122}$$

The equation error $\boldsymbol{e}$ can also be seen as the difference between filtered input $u(k)$ and output $y(k)$, respectively (see Figure 6.20). The central concept of the least-squares (LS) algorithm is to minimize the equation error with respect to θ. Therefore, the criterion

$$J = \sum_{i=k}^{k+N} e^2(i) = e^T e \rightarrow \min. \tag{6.123}$$

is defined. Inserting 6.122 in the criterion 6.123 yields

$$J = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) = \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} \rightarrow \min. \tag{6.124}$$

Differentiating 6.124 with respect to the parameter estimate $\hat{\theta}$ and setting equal to zero

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{\theta}}\bigg|_{\theta=\hat{\theta}} = -2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\theta}} = 0 \tag{6.125}$$

results in an equation for the parameter estimate. The LS estimate is therefore defined as

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \cdot \boldsymbol{y}. \tag{6.126}$$

The expression $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is called pseudo-inverse or Moore–Penrose inverse of $\boldsymbol{X}$. The non-singularity of $\boldsymbol{X}^T\boldsymbol{X}$ and the existence of the pseudo-inverse is guaranteed if the input signal is PE [6,13]. From Figure 6.20, the z-transform of the equation error $E(z)$ follows as

$$E(z) = -\boldsymbol{B}z^{-d}U(z) + \boldsymbol{A}Y(z) = -\boldsymbol{B}z^{-d}U(z) + \frac{\boldsymbol{A}}{\boldsymbol{A}_P}V(z) + \boldsymbol{A}\frac{\boldsymbol{B}_P z^{-d}}{\boldsymbol{A}_P}U(z). \tag{6.127}$$

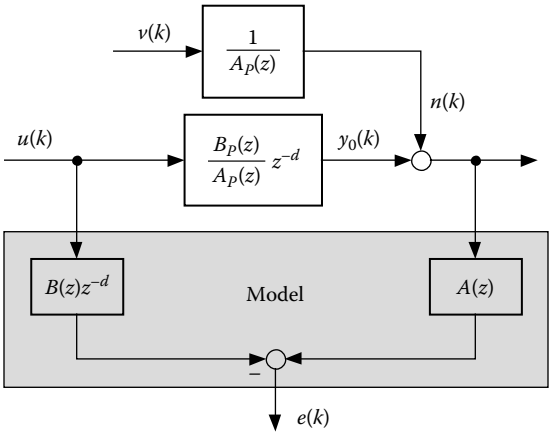


FIGURE 6.20 Noise filter for LS method (ARX model): measurement noise is generated by white noise filtered through a noise filter $1/A_p(z-1)$.

The identity $E(z) = V(z)$ in Equation 6.127 is fulfilled only if

$$\hat{\theta} = \theta, \quad \boldsymbol{A} = \boldsymbol{A}_P, \quad \boldsymbol{B} = \boldsymbol{B}_P \tag{6.128}$$

holds. Only in this case, the equation error is independent of the parameters and consists of a purely random signal (white noise).

The necessary assumptions for a LS estimator are as follows [13]:

- The LS method identifies an ARX model (see Equation 6.129), that is, a noise filter $\frac{1}{\boldsymbol{A}_p(z-1)}$ is assumed.
- The system order n (and dead time d) must be known.
- The input signal $u(k)$ must be known exactly.
- The input signal $u(k)$ must be PE of order $2n$.
- The noise signal $n(k)$ must be stationary with zero mean.

If these assumptions are fulfilled, the following statements will hold [13,14]:

1. The estimate is *unbiased*, that is, the mean estimation error is zero.
2. The estimate is *consistent*, that is, the parameter variance converges to zero if the number of samples approaches infinity.

In practical applications, the assumptions are verified by the model validation.

6.5.3.2.1 Model Structures

Apparently, the estimation of a noise filter is necessary in order to yield unbiased estimates. The following model structures differ only in their noise filter [13,14].

The *ARX model* has already been introduced:

$$Y(z) = \frac{B(z^{-1})}{A(z^{-1})}U(z) + \frac{1}{A(z^{-1})}V(z). \tag{6.129}$$

Only two polynomials A and B have to be estimated using the LS estimator. The ARX model is efficient and robust; but, since the noise transfer function is determined by the plant's denominator, it is not flexible for different noise spectra. In this case, it will yield biased estimates or a high model order.

The simplest structure is called *output-error model* (*OE model*), where the transfer function of the noise filter is 1.

$$Y(z) = \frac{B(z^{-1})}{A(z^{-1})}U(z) + \frac{1}{A(z^{-1})}V(z). \tag{6.130}$$

Although this structure looks even simpler than the ARX model, the parameters of the polynomials must be estimated using nonlinear prediction error methods (PEM).

The *Auto Regression with Moving Average and eXtra input* (ARMAX) *model* has a common denominator but a dedicated numerator for the noise filter:

$$Y(z) = \frac{B(z-1)}{A(z-1)}U(z) + \frac{D(z^{-1})}{A(z^{-1})}V(z). \tag{6.131}$$

The number of polynomials to estimate is three, but linear estimation methods may be employed. This model is the extension of the ARX model with added flexibility due to the numerator of the noise filter.

The most general model structure is the *Box–Jenkins model* (*BJ model*) with completely independent transfer functions for plant and noise filter:

$$Y(z) = \frac{B(z-1)}{A(z-1)}U(z) + \frac{D(z^{-1})}{A(z^{-1})}V(z). \tag{6.132}$$

It is capable of representing any dynamics and noise spectra, respectively, but with the highest computational effort.

6.5.3.3 Recursive Least-Squares Algorithm

In Equation 6.126, the parameter estimate is computed in a single step from all measured data. How can this estimate be improved when an additional measurement becomes available? Apparently, it would be very inefficient to repeat the computation with an augmented regressor matrix. Instead, a recursive formulation is sought for, where the current estimate $\hat{\boldsymbol{\theta}}(k+1)$ is computed using the previous estimate $\hat{\boldsymbol{\theta}}(k)$.

A derivation of recursive least squares (RLS) can be found in [6,13]. The resulting formulation is

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \boldsymbol{P}(k+1)x(k)\cdot[y(k+1) - \boldsymbol{x}^T(k+1)\hat{\boldsymbol{\theta}}(k)], \tag{6.133}$$

where $\boldsymbol{P}$ is defined as

$$\boldsymbol{P}(k) = (\boldsymbol{X}^T(k)\boldsymbol{X}(k))^{-1}. \tag{6.134}$$

Note that $\boldsymbol{P}$ is proportional to a covariance matrix of the parameter estimates (see Section 6.5.3.3). Equation 6.133 can be read as

$$\begin{aligned}\text{Current estimate} = {} & \text{previous estimate} + \text{correction vector} \\ & \times\big[\text{current measurement} - \text{prediction of the new measurement}\big].\end{aligned} \tag{6.135}$$

In this interpretation, the expression $\boldsymbol{x}^T(k+1)\hat{\boldsymbol{\theta}}(k)$ is the prediction of the new output. Although the current measurement $\boldsymbol{x}(k+1)$ is already available, the previous estimate $\hat{\boldsymbol{\theta}}(k)$ must be used for computation. This estimate is called a priori prediction (see Section 6.7.1.6).

The computation of 6.133 is problematic, since it requires a matrix inversion at each time step. Using a matrix inversion lemma [13], the correction vector $\gamma(k)$ can be computed according to

$$\gamma(k) = \frac{\boldsymbol{P}(k)\boldsymbol{x}(k+1)}{1 + \boldsymbol{x}^T(k+1)\boldsymbol{P}(k)\boldsymbol{x}(k+1)} \tag{6.136}$$

and the current parameter estimate is given by

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \gamma(k)[y(k+1) - \boldsymbol{x}^T(k+1)\hat{\boldsymbol{\theta}}(k)], \tag{6.137}$$

with $\boldsymbol{P}(k+1)$ recursively defined by

$$\boldsymbol{P}(k+1) = [\boldsymbol{I} - \gamma(k)\boldsymbol{x}^T(k+1)]\boldsymbol{P}(k). \tag{6.138}$$

Evaluating Equations 6.136 through 6.138 in this order allows for an efficient computation of the online estimate.

Example 6.14

The output of a process with discrete-time transfer function of

$$G(z^{-1}) = \frac{0.1^{z-1}}{1-0.9z^{-1}+0.5^{z-2}} \tag{6.139}$$

is corrupted by measurement noise (ARX structure). The input is white noise with a variance of $\sigma^2 u = 1$*, the noise source has a variance of* $\sigma^2 v = 0.14$*.*

The initial values for an RLS identification are $\boldsymbol{\theta}_0 = [0\ 0\ 0]'$ *and the main diagonal of* $\boldsymbol{P}_0$ *is* $10^6[1\ 1\ 1]'$*. After 150 samples, the estimated parameter values are* $\hat{\boldsymbol{\theta}} = [-0.95\ 0.52\ 0.13]'$ *and the corresponding entries of the* $\boldsymbol{P}$ *matrix are [0.027 0.027 0.0064]. The development of the parameter estimates as well as the diagonal entries of the* $\boldsymbol{P}$ *matrix can be seen in* Figures 6.21a,b*, respectively.*

Choice of initial values $\hat{\boldsymbol{\theta}}(0)$ *and* $\boldsymbol{P}(0)$

There are three possibilities for the choice of initial values [6,13]:

- Using an a priori estimate (rarely possible)
- Computing the LS estimate $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ and $\boldsymbol{P} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$ from the first m data and starting the RLS algorithm at $m + 1$
- Choosing a random initial parameter vector $\hat{\boldsymbol{\theta}}(0)$ and $\boldsymbol{P}(0) = \alpha\boldsymbol{I}$, where α is a large scalar

The choice of $P(0) = \alpha\boldsymbol{I}$, $\alpha \gg 1$ (reflecting an uncertain initial estimate) can be motivated as follows. Since

$$\hat{\boldsymbol{\theta}}(k) = \boldsymbol{P}(k)[\boldsymbol{X}^T(k)\boldsymbol{y}(k) + \boldsymbol{P}^{-1}(0)\hat{\boldsymbol{\theta}}(0)] \tag{6.140}$$

is valid for RLS, for $\alpha \to \infty$ and arbitrary $\hat{\boldsymbol{\theta}}(0)$ the LS estimator (Equation 6.126) will result.

The covariance of the parameter estimation error $cov(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is defined by [13]

$$cov(\boldsymbol{\theta} - \boldsymbol{\theta}) = \sigma_e^2\boldsymbol{P}, \tag{6.141}$$

where σ_e^2 is the variance of the residual. A covariance matrix $\boldsymbol{P}$ with large entries in the diagonal reflects the lack of information on the true parameter values.

6.5.3.4 Algorithms for the ARMAX Model

The parameter estimation of an ARMAX model poses a more complex problem: Since the inputs to the noise transfer function (white noise) are not known, they have to be reconstructed from measured data and current parameter estimates. Utilizing a recursive formulation, the parameters of the plant transfer function, the numerator polynomial of the noise transfer function, and the sequence of white noise entering the noise filter are estimated. Another method is the instrumental variable (IV) method [13].

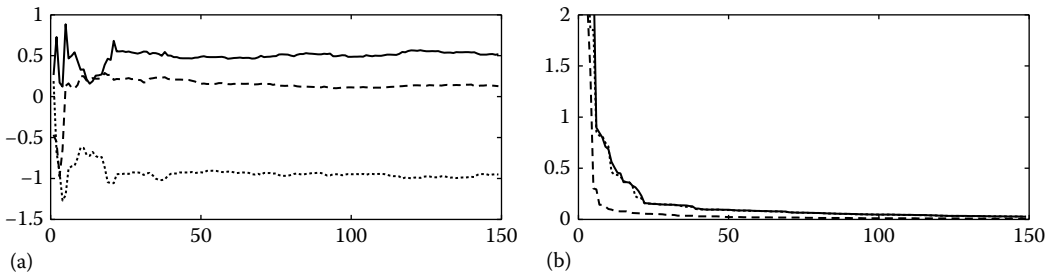


FIGURE 6.21 RLS identification. (a) The parameters show fast convergence from their initial values to the true ones. (b) The corresponding diagonal entries of the $\boldsymbol{P}$ matrix decrease rapidly.

6.5.3.5 RLS for Time-Varying Systems

In time-varying systems, the inherent memory of the parameter estimates should be limited in order to allow unbiased adaptation to new parameter values. A straightforward way to achieve this is given by applying a decaying exponential weight to the data such that the current measurement is most important and older measurements are neglected (weighted recursive least squares [WRLS]) [6,13]. The criterion for minimization J_k is defined as

$$J_k = \sum_{i=k}^{k} \lambda^{k-i} e^2(i) = e_k{}^T e_k, \quad 0 < \lambda < 1, \tag{6.142}$$

where λ is the so-called *forgetting factor*. Ordinary RLS is obtained with $\lambda = 1$.

The resulting recursive estimation algorithm consists of three formulas:

$$\gamma(k) = \frac{1}{\boldsymbol{x}^T(k+1)\boldsymbol{P}(k)\boldsymbol{x}(k+1) + \lambda} \boldsymbol{P}(k)\boldsymbol{x}(k+1), \tag{6.143}$$

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \gamma(k)\,[y(k+1) - \boldsymbol{x}^T(k+1)\hat{\theta}(k)], \tag{6.144}$$

$$\boldsymbol{P}(k+1) = [\boldsymbol{I} - \gamma(k)\boldsymbol{x}^T(k+1)]\boldsymbol{P}(k)\frac{1}{\lambda}. \tag{6.145}$$

Choice of the forgetting factor λ

- Choose λ small (e.g., $\lambda = 0, 9$), if the parameters change quickly. In this case, measurement noise strongly affects the estimates (large estimation error covariance).
- Choose λ large (e.g., $\lambda = 0, 999$), if the parameters change slowly. Measurement noise is attenuated (small estimation error covariance).

A time-varying forgetting factor can also be utilized to achieve a good estimate [6,13].

Example 6.15

The forgetting factor λ acts as a low-pass filter. For values of λ close to 1, a smooth but slow response to a stepwise parameter change follows. For smaller λ, the variance increases but the estimate reacts more quickly (see Figure 6.22). For $\lambda = 1$, a biased estimate results, since the measurements before the parameter change corrupt the estimate.

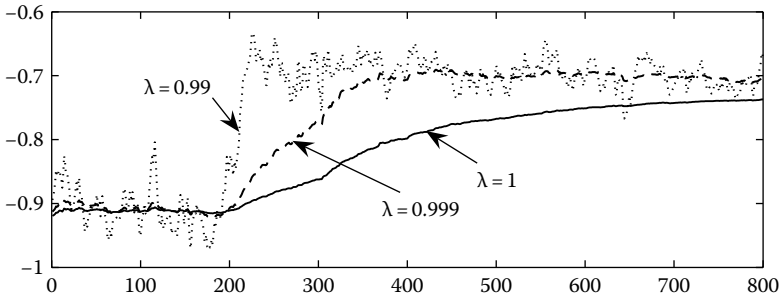


FIGURE 6.22 WRLS identification: for $\lambda = 0.999$ a compromise can be achieved, $\lambda = 0.99$ shows large variance, and $\lambda = 1$ is far too slow in reaching the new value. In this case, a biased estimate results.

6.5.3.6 Prediction Error Method

In the case of ARX and ARMAX model, the parameter estimate $\hat{\boldsymbol{\theta}}$ can be computed using linear methods. However, BJ and OE model require nonlinear estimation methods, since minimization of a quadratic criterion using the prediction error $\varepsilon(k|\boldsymbol{\theta}) = y(k) - \hat{y}(k|\boldsymbol{\theta})$ leads to a nonlinear optimization problem for the parameter vector $\hat{\boldsymbol{\theta}}$. Therefore, general optimization methods like Gauss–Newton or Levenberg–Marquard have to be employed. The drawback of these methods is their dependency on initial conditions, optimization parameters and possibly convergence problems [13].

6.5.3.7 Identification of a State-Space Model

Identification of a state-space model (6.1, 6.2) requires estimation of a higher number of parameters, and the additional problem of generating a nonunique sequence of state vectors has to be solved. Nevertheless, algorithms for direct identification of MIMO state-space systems exist [13,15] and can be applied especially to systems with high number of inputs and outputs.

6.5.4 Black-Box versus Gray-Box Models

Until now, the model structure consists of the assumption of linearity, the model orders (dead time), and the noise filter. Usually, these features have to be chosen a priori and after an appropriate identification step the validation results indicate whether these assumptions were justified or not. Since no information is available at the beginning of the identification, these models are called black-box models.

In the case of physical models, however, the model structure is already known due to analytical or empirical information and only the parameter values of this model have to be found. To indicate that some structural information already exists, these models are called gray-box models.

Example 6.16

The relation between spring force F and compression s in a linear spring is given by the expression $F(s) = ks$, where k is the linear spring constant (measured, e.g., in N/mm). A nonlinear spring characteristic is often desired in automotive suspensions. Typical equation for such a progressive spring is

$$F(s) = ks + hs^3, \tag{6.146}$$

where h is an additional coefficient in N/mm^3. In Figure 6.23, the original characteristic with $k = 50$ N/mm and $h = 0.0015$ N/mm^3 is plotted as solid line. Note that this model is nonlinear but linear in the parameters.

Assuming that 25 data (s_i, F_i) have been measured (circles in Figure 6.23), the parameters k *and* h will be identified. By building a regression vector $[s_i \quad s_i^3]$ together with the corresponding output value F_i, the least-squares solution 6.126 can be computed leading to the estimates $\hat{k} = 59.3$ N/mm and $\hat{h} = 0.0010$ N/mm^3. The model prediction is plotted as dashed line in Figure 6.23. The identified model lies close to the true one despite the small number of measurements. Observe that the identified model starts exactly in the origin since this is required by the model structure.

6.5.5 Nonlinear Modeling

Nonlinear models provide an abundance of structures, and the choice of a distinct structure is always motivated by the individual application. Only a short and incomplete introduction can be given here.

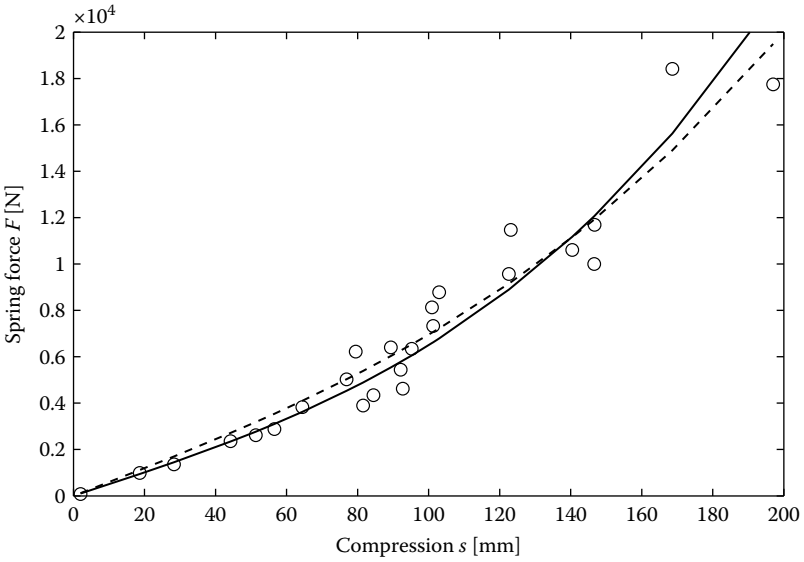


FIGURE 6.23 Gray-box identification of a progressive spring. Measured data (circles), true model (solid line), and identified model (dashed line).

6.5.5.1 Models that are Linear in the Parameters

This important class of models can be described by

$$y_i = f_1(\boldsymbol{y}, \boldsymbol{u})\theta_1 + \cdots + f_n(\boldsymbol{y}, \boldsymbol{u})\theta_n = [f_1(\boldsymbol{y}, \boldsymbol{u}) \cdots f_n(\boldsymbol{y}, \boldsymbol{u})]\boldsymbol{\theta}, \tag{6.147}$$

where

y_i is an output value
$\boldsymbol{u}$ and $\boldsymbol{y}$ are the input and output vectors, respectively
$\boldsymbol{\theta}$ is the parameter vector
f_i are n nonlinear functions on the input and output vectors, respectively

In the case of vanishing $\boldsymbol{y}$ in the arguments of the f_i, the output directly depends on the inputs and a moving average (MA) model results. In the case of past inputs and outputs in the arguments of the f_i, a nonlinear autoregressive model with moving average and exogenous input (NARMAX) is given. It is important to notice that 6.147 is of the same structure as 6.119; hence, a regressor matrix with nonlinear function values can be constructed and the pseudo-inverse (6.126) can be applied [13].

Polynomial models or cubic splines exploit this structure, which combines computational efficiency with extended flexibility compared to linear models.

6.5.5.2 Local Linear Models

The quality of an approximation of a global nonlinear function using local linear models critically depends on the regions of validity for each local model. The main focus of these methods is therefore the partitioning of the input space, which is a nonlinear optimization problem. Several model structures like LOLIMOT, INN, RBF have been proposed [16] and are state of the art in automotive applications like ECU mappings.

6.5.5.3 Artificial Neural Networks

An artificial neural network (ANN) is a nonlinear structure that is capable of approximating arbitrary nonlinearities; but, since the solution is not unique, many different network structures are possible. On the one hand, the choice of the network topology, the activation functions, and the

(nonlinear) training algorithm constitute a flexible tool, which is sometimes capable of representing a complex nonlinear behavior in compact formulation. On the other hand, it requires expert knowledge to efficiently utilize these tools and not get lost in the numerous possibilities [7].

6.5.6 Stepwise Identification Procedure

6.5.6.1 Gathering Information

The first and most important step to an experimental model is measuring the inputs and outputs of the process. Defining the system borders, keeping in mind the purpose of the model, and choosing appropriate excitation signals are of fundamental importance to any identification. The most important issues for a successful identification are listed as follows [14]:

- The system borders should be chosen such that the complexity of the unknown model is minimal; all significant inputs and outputs can be measured; and the dynamics of interest are contained.
- Data recording must be performed according to correct measurement procedures, sampling frequency and any filters must be chosen to match the expected model characteristic.
- The input signals must fulfill the condition of being PE (optimal: white noise).
- The input signals must feed enough energy into the system in order to achieve an acceptable signal-to-noise ratio in the outputs.
- The input signals must be generated according to the operating point of the process. This could include consecutive stepwise changes in order to cover a wide dynamic range (possibly nonlinear) with local models. Also ramps can be used.

Careful planning resulting in a measuring protocol and a complete and comprehensive documentation is essential.

6.5.6.2 Pre-Processing of Data

All data must be plotted at least in the time domain in order to recognize outliers, artefacts, offsets, drifts, and significant changes in the system dynamics. Many of these defects can be filtered automatically; however, manual control of these procedures is highly recommended. In the case of much higher sampling rates than necessary for the system dynamics, a correct downsampling improves the efficiency of the algorithms.

All data should be split into a larger identification part and a smaller validation part (necessary for cross-validation, see Section 6.4.3). Care should be taken to have the same energy and frequency content in both data sets.

6.5.6.3 Nonparametric Models

If no correlation between measurement noise and input exists (as in closed loop), the impulse response and the frequency response are important tools. Both provide the first information on the absolute and relative degree of the model, time constants, stationary gain, and damping. The impulse response estimate is especially useful for estimating the dead time of a system. Using confidence intervals for the zero line, a quantitative conclusion on the probability of a certain dead time is possible (Figure 6.24a).

6.5.6.4 Parametric models

Parametric linear models should be estimated beginning with the ARX structure and ending with the most complex BJ model. For the purpose of dead-time estimation, a low-order (third- or fourth-order) ARX model can be identified for different dead times. The model with the correct dead time usually yields a significantly better prediction performance than the others. Once the dead time has been ascertained, it is fixed for all consecutive models.

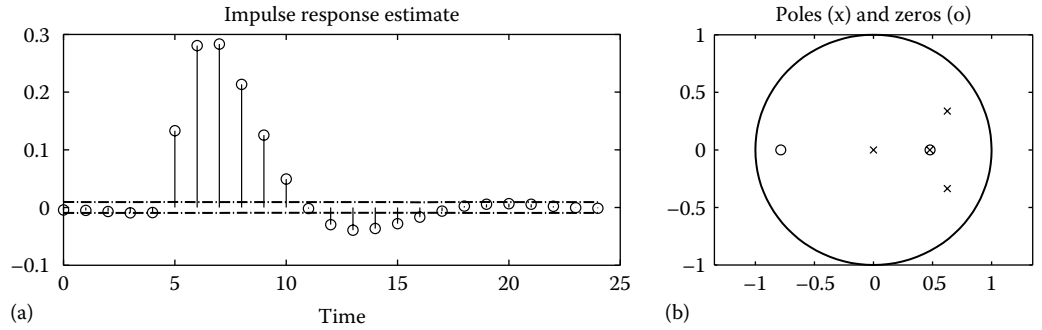


FIGURE 6.24 (a) Estimate of the impulse response by a correlation analysis; a dead time of five samples is observable. (b) Pole-zero plot; a cancellation occurs at (0.5/0).

Utilizing the validation methods outlined later, the model orders for numerator and denominator polynomials of plant and noise model are increased in an iterative manner. For each model structure, the model with the best trade-off between complexity and performance is saved, and finally, the overall best linear model is chosen between them.

6.5.6.5 Model Validation

The following procedures have to be carried out for each model and form the basis of a fair comparison in order to find the optimal model.

6.5.6.5.1 Model Performance

By means of a numerical simulation, the output of a certain model is computed from the validation input. This predicted output is compared to the measured validation output (model fit). In general, a good model will have a good fit; however, due to the measurement noise a perfect fit is not desirable. Models with very high order tend to reproduce the measurement noise in the identification data, but due to cross-validation they will yield a smaller performance on the validation data compared to a lower-order model.

6.5.6.5.2 Residual Analysis

The residual analysis computes the ACF and CCF between the residual and the input, respectively. The residual due to measurement noise should be purely stochastic and therefore not correlated with itself or the input signal [13]. Its autocorrelation should therefore be a Kronecker-delta function.

The correlation functions are sensitive tools for detecting significant correlations by showing signal values significantly different from zero as indicated by 99% confidence intervals (Figure 6.25). In this case, either the model order or the model structure is not correct, or the general assumptions are not fulfilled (e.g., by nonlinear dynamics or closed-loop data).

6.5.6.5.3 Model Order

The optimal model order is found applying the procedure outlined earlier. For too low orders, both performance and residual analysis will show poor results. Right above the optimal model order, the performance will hardly increase or even decrease. The residual analysis will not detect too high model orders, but a simple pole-zero plot can detect near cancellations, which are caused by redundant poles and zeros (Figure 6.24b). Small models typically show a better ability to generalize, that is, to achieve a good model fit on data not available for identification.

6.5.6.5.4 Model Structure

In order to have unbiased estimates, the correct noise filter has to be found. The most important tool for this purpose is again the correlation analysis. If a wrong model structure has been chosen,

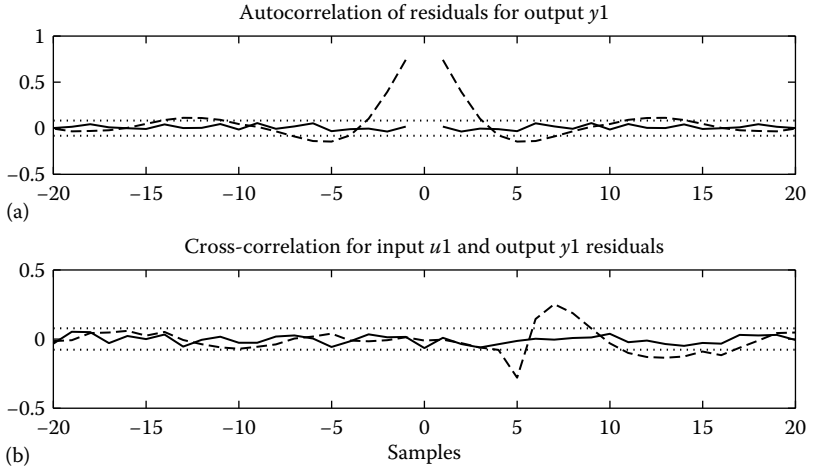


FIGURE 6.25 Residual analysis: (a) ACF of the residual and (b) CCF between residual and input show satisfactory results for an ARX model (solid line) while a low-order OE model (dashed line) exceeds the 99% confidence intervals.

a satisfactory result is only possible for very high orders. Furthermore, an adequate ACF of the residual but significant cross-correlations between input and residual indicate a wrong (too low) order of the numerator, and vice versa.

As opposed to linear system theory, nonlinear control concepts cannot be covered by a general approach. The theory covering nonlinear systems is also complex and mathematically involved. Textbooks covering the general concepts and may be used for an introduction are [8,9,11]. Therefore, the optimal strategy is to find a matching design for a specific nonlinear control problem. Following is a listing of some important techniques [9], which may be applied to particular control problems:

Trial and error: Using nonlinear analysis methods and considering the special structure of a given control problem, a custom-tailored control concept is designed and verified by simulations. If the complexity of the problem becomes too large, this approach will most likely fail.

Linearization: A straightforward method for applying the well-known and powerful linear design tools is to linearize a given nonlinear system model around an operating point. Obviously, the system states are required to stay within a vicinity of this operating point; otherwise, the linearization may no longer be valid. In this case, poor closed-loop performance and even unstable system behavior can result. Nevertheless, since this approach utilizes linear controller design methods, it is widely used and therefore elaborated on in the remainder of this chapter.

Gain scheduling: Gain scheduling is the extension to linearization around only one operating point by selecting a number of operating points which cover the whole operating range. By computing a linear controller for each operating point, discrete sets of controller parameters will result. Between operating points, the controller parameters will be interpolated or scheduled in order to have a smooth change in parameters.

Robust control: Most model-based design technique assumes a valid nominal model and model uncertainties are not explicitly incorporated into the design process. In robust control, these uncertainties are quantitatively considered during the controller design and a guaranteed performance is obtained as long as the assumptions on the uncertainties apply. A linearized model is often the starting point for a robust controller design, and the nonlinearities are then regarded as uncertainties of the nominal linear model.

Adaptive control: Another approach to deal with nonlinear or time-varying systems is adaptive control. Basically, the dynamic structure of the system has to be known but the unknown or slowly varying parameters are identified online. The resulting overall control system is inherently nonlinear.

6.6 NONLINEAR CONTROL DESIGN

6.6.1 Linearization

A time-invariant MIMO system is generally described by a nonlinear state-space differential equation and an associated output equation:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{z}); \qquad \boldsymbol{x}(t_o) = \boldsymbol{x}_o \tag{6.148}$$

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{z}) \tag{6.149}$$

where

$\boldsymbol{x}$ der $(n \times 1)$ is the state vector
$\boldsymbol{u}$ der $(r \times 1)$ is the input vector
$\boldsymbol{z}$ der $(q \times 1)$ is a vector of disturbance inputs
$\boldsymbol{y}$ der $(m \times 1)$ is the vector of outputs

6.6.1.1 Linearization about an Equilibrium Point

An equilibrium point $\boldsymbol{x}_R$ of the system (6.148) can be determined by setting the state derivative to zero:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}_R, \boldsymbol{u}_R, \boldsymbol{z}_R) = \boldsymbol{0}. \tag{6.150}$$

The equilibrium output $\boldsymbol{y}_R$ is determined from

$$\boldsymbol{y}_R = \boldsymbol{g}(\boldsymbol{x}_R, \boldsymbol{u}_R, \boldsymbol{z}_R). \tag{6.151}$$

The linearized dynamics of 6.150 about $\boldsymbol{x}_R$ are then described by

$$\Delta\dot{\boldsymbol{x}} = A\Delta\boldsymbol{x} + B\Delta\boldsymbol{u} + E\Delta\boldsymbol{z} \tag{6.152}$$

and

$$\Delta y = C\Delta x + D\Delta u + F\Delta z. \tag{6.153}$$

Here, $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_R$, $\Delta\boldsymbol{u} = \boldsymbol{u} - \boldsymbol{u}_R$, $\Delta\boldsymbol{z} = \boldsymbol{z} - \boldsymbol{z}_R$ and $\Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}_R$ are used to describe deviations from the equilibrium point.

The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, and $\boldsymbol{F}$ are obtained from

$$\boldsymbol{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}_R \quad \boldsymbol{B} = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_r} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdots & \frac{\partial f_2}{\partial u_r} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \frac{\partial f_n}{\partial u_2} & \cdots & \frac{\partial f_n}{\partial u_r} \end{bmatrix}_R$$

$$\boldsymbol{E}=\left[\begin{array}{cccc}\frac{\partial f_1}{\partial z_1} & \frac{\partial f_1}{\partial z_2} & \cdots & \frac{\partial f_1}{\partial z_q}\\ \frac{\partial f_2}{\partial z_1} & \frac{\partial f_2}{\partial z_2} & \cdots & \frac{\partial f_2}{\partial z_q}\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \frac{\partial f_n}{\partial z_1} & \frac{\partial f_n}{\partial z_2} & \cdots & \frac{\partial f_n}{\partial z_q}\end{array}\right]_R \quad \boldsymbol{C}=\left[\begin{array}{cccc}\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n}\\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n}\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n}\end{array}\right]_R$$

$$\boldsymbol{D}=\left[\begin{array}{cccc}\frac{\partial g_1}{\partial u_1} & \frac{\partial g_1}{\partial u_2} & \cdots & \frac{\partial g_1}{\partial u_r}\\ \frac{\partial g_2}{\partial u_1} & \frac{\partial g_2}{\partial u_2} & \cdots & \frac{\partial g_2}{\partial u_r}\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \frac{\partial g_m}{\partial u_1} & \frac{\partial g_m}{\partial u_2} & \cdots & \frac{\partial g_m}{\partial u_r}\end{array}\right]_R \quad \boldsymbol{F}=\left[\begin{array}{cccc}\frac{\partial g_1}{\partial z_1} & \frac{\partial g_1}{\partial z_2} & \cdots & \frac{\partial g_1}{\partial z_q}\\ \frac{\partial g_2}{\partial z_1} & \frac{\partial g_2}{\partial z_2} & \cdots & \frac{\partial g_2}{\partial z_q}\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \frac{\partial g_m}{\partial z_1} & \frac{\partial g_m}{\partial z_2} & \cdots & \frac{\partial g_m}{\partial z_q}\end{array}\right]_R$$

or in short notation

$$\boldsymbol{A}=\left.\frac{\partial \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{x}}\right|_R \qquad B=\left.\frac{\partial \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{u}}\right|_R$$

$$E=\left.\frac{\partial \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{z}}\right|_R \qquad C=\left.\frac{\partial \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{x}}\right|_R$$

$$D=\left.\frac{\partial \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{u}}\right|_R \qquad F=\left.\frac{\partial \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u},\boldsymbol{z})}{\partial \boldsymbol{z}}\right|_R .$$

6.6.2 Feedback Linearization

Using control authority to transform nonlinear models into linear ones is one of the most commonly used ideas of practical nonlinear control design. Feedback linearization is a versatile control technique for nonlinear systems, which converts a nonlinear system into a linear system using a nonlinear coordinate transformation and feedback control.

It is noteworthy that feedback linearization relies on access to accurate information about the control system in terms of its parameters. While in some cases one can extend the benefits of feedback linearization to approximately known and imperfectly observed models, information flow constraints remain a serious obstacle when applying feedback linearization.

For the sake of simplicity, the explanation is restricted to nonlinear time-invariant SISO systems of the form:

$$\dot{\boldsymbol{x}}=\boldsymbol{f}(\boldsymbol{x}(t))+\boldsymbol{g}(\boldsymbol{x}(t))u(t) \quad \boldsymbol{x}\in\mathbb{R}^n, \tag{6.154}$$

$$y(t) = h(x(t)). \tag{6.155}$$

Feedback linearization now aims at finding a (generally nonlinear) state-space transformation $z = \Phi(x)$ and a feedback law that transforms the original nonlinear system (6.154) into a linear one so that linear control theory can be applied subsequently.

6.6.2.1 Zero Dynamics—Output Zeroing

Definition 6.4

Zero dynamics: The term zero dynamics is used to describe the internal dynamics of a system when the output y(t) is identical zero for a certain input signal u(t) and initial conditions $\boldsymbol{x}$*(0).*

For the computation of the zero dynamics, initial conditions and an input signal have to be found such that

$$y(t) = \dot{y}(t) = \ddot{y}(t) = \cdots = 0 \tag{6.156}$$

holds for all $t \geq 0$. This immediately leads to the first requirement:

$$y(t) = h(x(t)) = 0. \tag{6.157}$$

The requirement (6.157) is independent of the system input $u(t)$ and can thus only be fulfilled by proper selection of the initial conditions.

Differentiating the output equation (6.155) yields

$$\dot{y}(t) = \left[\frac{\partial h(\boldsymbol{x}(t))}{\partial \boldsymbol{x}}\right]^T \dot{\boldsymbol{x}}(t) = \left[\frac{\partial h(\boldsymbol{x}(t))}{\partial \boldsymbol{x}}\right]^T [f(\boldsymbol{x}(t)) + \boldsymbol{g}(\boldsymbol{x}(t))u(t)] \tag{6.158}$$

and setting $\dot{y}(t) = 0$ yields

$$\dot{y}(t) = L_f h(\boldsymbol{x}(t)) + L_g h(\boldsymbol{x}(t))u(t) = 0. \tag{6.159}$$

The expressions $L_f h(\boldsymbol{x}(t))$ and $L_g h(\boldsymbol{x}(t))$ are the so-called *Lie derivatives* or directional derivatives of $h(\boldsymbol{x})$ along $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$, respectively.

If $L_g h(\boldsymbol{x}(t)) = 0$ in 6.159, then $\dot{y}$ is also independent of $u(t)$ and one proceeds with the second derivative:

$$\ddot{y}(t) = \frac{\partial y(t)}{\partial t} = \left[\frac{\partial L_f h(\boldsymbol{x}(t))}{\partial \boldsymbol{x}}\right]^T \dot{\boldsymbol{x}}(t) = \left[\frac{\partial L_f h(\boldsymbol{x}(t))}{\partial \boldsymbol{x}}\right]^T [\boldsymbol{f}(\boldsymbol{x}(t)) + \boldsymbol{g}(\boldsymbol{x}(t))u(t)] \tag{6.160}$$

and setting $\ddot{y}(t) = 0$ yields

$$\ddot{y}(t) = L_f^2 h(\boldsymbol{x}(t)) + L_g L_f h(\boldsymbol{x}(t))u(t) = 0. \tag{6.161}$$

If $L_g L_f h(\boldsymbol{x}(t)) = 0$, then $u(t)$ has nondirect influence on the second derivative of the output as well. In this case, the third derivative of the output is obtained as

$$y^{(3)}(t) = L_f^3 h(\boldsymbol{x}(t)) + L_g L_f^2 h(\boldsymbol{x}(t))u(t). \tag{6.162}$$

Derivatives of the output are computed consecutively until the input $u(t)$ explicitly appears in the rth derivative $y^{(r)}(t)$:

$$L_g L_f^{r-1} h(\boldsymbol{x}(t)) \neq 0 \tag{6.163}$$

which implies that

$$y^{(r)}(t) = L_f^r h(\boldsymbol{x}(t)) + L_g L_f^{r-1} h(\boldsymbol{x}(t)) u(t). \tag{6.164}$$

The number r is then called the *relative degree* of the system. Zeroing the output $y(t)$ at all times $t \geq 0$ according to 6.156 implies that the initial conditions $\boldsymbol{x}(t = 0)$ have to be selected such that

$$h(\boldsymbol{x}) = L_f h(\boldsymbol{x}(t)) = L_f^2 h(\boldsymbol{x}(t)) = \cdots = L_f^{r-1} h(\boldsymbol{x}(t)) = 0 \tag{6.165}$$

holds.

A suitable state-space transformation for feedback linearization is now obtained through

$$\begin{aligned} z_1 &= \Phi_1(\boldsymbol{x}) = h(\boldsymbol{x}) \\ z_2 &= \Phi_2(\boldsymbol{x}) = L_f h(\boldsymbol{x}(t) \\ &\vdots \quad \vdots \\ z_r &= \Phi_r(\boldsymbol{x}) = L_f^{r-1} h(\boldsymbol{x}(t)). \end{aligned} \tag{6.166}$$

The remaining transformations $\Phi_{r+1}(\boldsymbol{x})$ through $\Phi_n(\boldsymbol{x})$ have to be found such that the Jacobian $\partial\Phi/\partial x \neq 0$ and that $L_g \Phi_i(\boldsymbol{x}) = 0$ holds.

In the new coordinates, the first r equations of the state-space system now read

$$\begin{aligned} \dot{z}_1 &= z_2 \\ \dot{z}_2 &= z_3 \\ &\vdots \quad \vdots \\ \dot{z}_{\mathrm{r}-1} &= z_r \\ \dot{z}_r &= L_f^r h(\boldsymbol{x}(t)) + L_g L_f^{r-1} h(\boldsymbol{x}(t)) u(t) = a(\boldsymbol{x}) + b(\boldsymbol{x}) u. \end{aligned} \tag{6.167}$$

Next, the linearizing feedback law is defined as

$$u(t) = \frac{-a(\boldsymbol{x}) + v(t)}{b(\boldsymbol{x})}, \tag{6.168}$$

where $v(t)$ is the control input associated with the linear system. Application of $u(t)$ requires that $b(\boldsymbol{x}) \neq 0$. Inserting into the last derivative (6.167) leads to

$$\dot{z}_{\mathrm{r}} = v(t). \tag{6.169}$$

The structure of the feedback-linearized system is illustrated in Figure 6.26.

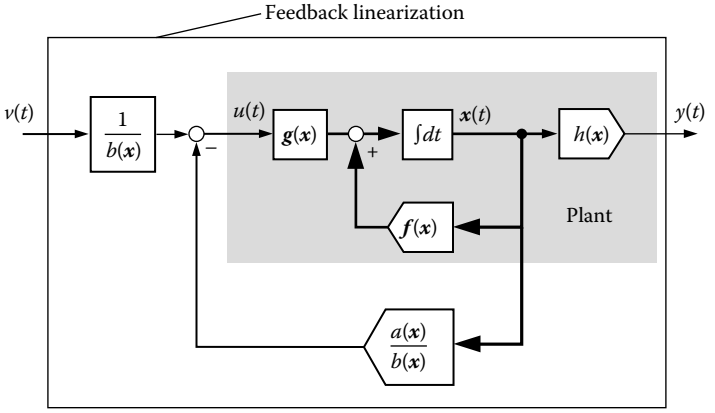


FIGURE 6.26 Block diagram of a feedback-linearized system.

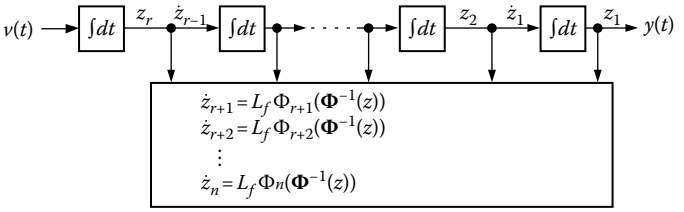


FIGURE 6.27 Resulting structure of a feedback-linearized system.

If the remaining transformations are chosen such that $L_g\Phi_i(\boldsymbol{x}) = 0$, then

$$\begin{aligned}
\dot{z}_{r+1} &= q_{r+1}(z) = \mathrm{L}_f\Phi_{r+1}(\boldsymbol{\Phi}^{-1}(z)) \\
\dot{z}_{r+2} &= q_{r+2}(z) = \mathrm{L}_f\Phi_{r+2}(\boldsymbol{\Phi}^{-1}(z)) \\
&\vdots \quad \vdots \\
\dot{z}_n &= q_n(z) = \mathrm{L}_f\Phi_n(\boldsymbol{\Phi}^{-1}(z)).
\end{aligned} \tag{6.170}$$

The transformed system now consists of two parts: first, there is a sequence of r integrators, described by 6.167 with $v(t)$ as input to the rth integrator according to 6.169. This subsystem is linear and fully controllable and observable. The second part, described by 6.170, is generally nonlinear and has no influence on the output $y(t)$. It is therefore also understood as the *unobservable* part of the transformed system. The situation is illustrated in Figure 6.27. If $v(t) \equiv 0$ and if all conditions (6.165) are met, then the unobservable part of the system (6.170) is commonly called *zero dynamics*.

For a subsequent controller design for the linear part $z_1, \ldots, z_r$, it is essential that the zero dynamics $z_{r+1}, \ldots, z_n$ are stable.

Example 6.17

Feedback linearization

Consider the system

$$\dot{x} = \begin{bmatrix} -x_1 \\ x_1x_2 \\ x_2 \end{bmatrix} + \begin{bmatrix} \exp(x_2) \\ 1 \\ 0 \end{bmatrix} u(t) \quad \text{and} \quad y = h(x) = x_3. \tag{6.171}$$

From 6.157, one gets

$$y = h(x) = x_3 = 0 \qquad z_1 = \Phi_1(x) = x_3.$$

Next, 6.159 is evaluated:

$$\dot{y} = L_f h(\boldsymbol{x}(t) + L_g h(\boldsymbol{x}(t) u(t) = x_2 = 0 \qquad z_2 = \Phi_2(\boldsymbol{x}) = x_2.$$

Then, 6.161 gives

$$\ddot{y} = L_f^2 h(\boldsymbol{x}(t)) + L_g L_f h(\boldsymbol{x}(t)) u(t) = x_1 x_2 + u = 0.$$

Apparently, the relative degree of the system is $r = 2$ and the control law (6.168) now reads

$$u(t) = -x_1 x_2 + v(t).$$

Next, $\Phi_3(\boldsymbol{x})$ has to be found such that

$$L_g \Phi_3(\boldsymbol{x}) = \left[\frac{\partial \Phi_3(\boldsymbol{x})}{\partial \boldsymbol{x}} \right]^T \boldsymbol{g}(\boldsymbol{x}) = \frac{\partial \Phi_3}{\partial x} \exp(x_2) + \frac{\partial \Phi_3}{\partial x_1} 1 = 0.$$

One possible solution is $\Phi_3(\boldsymbol{x}) = 1 + x_1 - \exp(x_2)$. The Jacobian then becomes

$$\frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{x}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & -\exp(x_2) & 0 \end{bmatrix}.$$

Accordingly, the feedback-linearized system is described by

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} z_2 \\ v(t) \\ (1 - z_3 - \exp(z_2))(1 + z_2 \exp(z_2)) \end{bmatrix}. \tag{6.172}$$

The zeroing initial conditions (6.165) become

$$h(x) = z_1 = x_3 = 0, \quad L_f h(\boldsymbol{x}(t)) = z_2 = x_2 = 0, \quad \text{and thus } u = 0$$

Inserting into the last equation (6.172) shows that the zero dynamics for $u = 0$ are described by $\dot{z}_3 = -z_3$. In this case, the zeros dynamics are linear and stable.

6.6.3 Robust Control

The robust control concept most widely used is H_∞ control. This is due to the seamless fit in the framework of optimal control and versatile design parameters which allow for the incorporation of different types of model uncertainties. In this section, only a short overview will be given and thorough coverage is given in [21].

The definition of the H_∞ norm is given by

$$\left\| \mathbf{G} \right\|_\infty = \sup_w \frac{\left\| \mathbf{y} \right\|_2^2}{\left\| \mathbf{u} \right\|_2^2}, \tag{6.173}$$

where $\mathbf{G}$ is the closed-loop transfer function between the input $\mathbf{u}$ and the output $\mathbf{y}$.

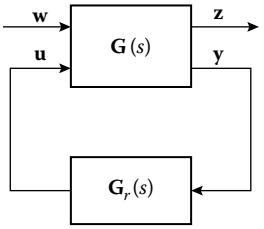


FIGURE 6.28 Basic structure of the H_∞-control problem.

The formulation of the H_∞-control problem is based on the structure in Figure 6.28. The system input is now **w** whereas the output is some arbitrary performance variable **z**. An obvious and important differentiation to LQR design is that disturbance and control inputs as well as measured and performance outputs need not be identical. They all will, however, be explicitly incorporated into the controller design. This fact enables the designer to achieve a predefined controller performance (defined only in the performance variable **z**) in spite of measuring other output variables **y**. Nevertheless, the fundamental principle behind H_∞ design is also an optimization procedure.

The transfer-function matrix **G** in Figure 6.28 can be partitioned into four appropriate submatrices $\mathbf{G}_{ij}$, which constitute the following equations for the variables **z** and **y**:

$$\mathbf{z} = \mathbf{G}_{zw}\mathbf{w} + \mathbf{G}_{zu}\mathbf{u}, \tag{6.174}$$

$$\mathbf{y} = \mathbf{G}_{yw}\mathbf{w} + \mathbf{G}_{yu}\,\mathbf{u}, \tag{6.175}$$

where the $\mathbf{G}_{ij}$ are the respective components of the plants transfer-function matrix. Combining Equations 6.174 and 6.175, the input/output characteristic of the plant **G** and of the controller $\mathbf{G}_r$ are obtained by

$$\begin{bmatrix}\mathbf{z}\\ \mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_{zw} & \mathbf{G}_{zu}\\ \mathbf{G}_{yw} & \mathbf{G}_{yu}\end{bmatrix}\begin{bmatrix}\mathbf{w}\\ \mathbf{u}\end{bmatrix} \tag{6.176}$$

and

$$\mathbf{u} = \mathbf{G}_r\mathbf{y}. \tag{6.177}$$

The transfer function of the overall closed-loop system is given by

$$\mathbf{T}_{zw} = \frac{\mathbf{z}}{\mathbf{w}} \tag{6.178}$$

and by inserting Equation 6.176 into Equation 6.175 it is transformed to

$$\mathbf{T}_{zw} = \mathbf{G}_{zw} + \mathbf{G}_{zu}\mathbf{G}_r(1-\mathbf{G}_{yu}\mathbf{G}_r)^{-1}\mathbf{G}_{yw}. \tag{6.179}$$

The result of 6.179 is termed as lower fractional transform (LFT). This description is unique and can be directly computed from Equation 6.176. Using Equation 6.179, the transfer zeros of the closed loop can be affected by an appropriate choice of $\mathbf{G}_{zw}$ and $\mathbf{G}_{zu}$. The minimization of the LFT with respect to the ∞ norm is the H_∞-design problem and can be recasted as

$$\min_{\mathbf{G}_r(s)} \left\|\mathbf{T}_{zw}\right\|_\infty = \min_{\mathbf{G}_r(s)} \sup_{w\in R} \bar{\sigma}\{\mathbf{T}_{zw}\}. \tag{6.180}$$

In the case of a SISO system, the minimization of the largest singular value $\bar{\sigma}$ is identical to minimizing the maximal value of the gain response of $\mathbf{T}_{zw}$. In contrast to LQ optimization, where the integral criterion averages over the frequency range, H_∞ design aims to limit the gain response only in the direction of the worst case.

Some advantages and disadvantages of the H_∞-control design method are outlined in the following:

Advantages:

- A controller design explicitly incorporating all disturbances is possible.
- An assumption of uncorrelated (and possibly white) disturbances and control inputs is not necessary.
- Arbitrary performance variables can be defined and the control performance is guaranteed.
- Frequency weighting of the inputs and outputs is easily implemented.
- The effects of neglected structural or actuator dynamics can be counteracted.
- Standard method for robust control, many numerical tools available.

Disadvantages:

- The design of the controller is mathematically involved and especially scaling issues of the plant model can be cumbersome.
- Inherent coupling of an observer and controller design ⇒ the optimization has to be solved iteratively. However, software tools considerably ease this situation.

6.7 STATE ESTIMATION REVISITED

The basic state observer has already been introduced in Section 6.2.2 (Luenberger observer). However, the Luenberger observer has been derived under the assumption of both deterministic inputs and perfect knowledge of the stationary plant. If these conditions are not met, the performance of the observer may be poor or it may even become unstable. In this chapter, two concepts are introduced, which are able to deal with the earlier outlined problems: the Kalman filter for systems with stochastic inputs, and the sliding mode observer (SMO) for systems with parameter variations.

6.7.1 Kalman Filter

6.7.1.1 Motivation

The Kalman filter problem is fundamentally different from the previous system configurations. Not only measurement noise $\boldsymbol{\nu}(t)$ corrupts the output measurement, but also an additional stochastic input called process noise $\boldsymbol{w}(t)$ is fed into the system (see also Figure 6.29):

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{G}\boldsymbol{w}; \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{6.181}$$

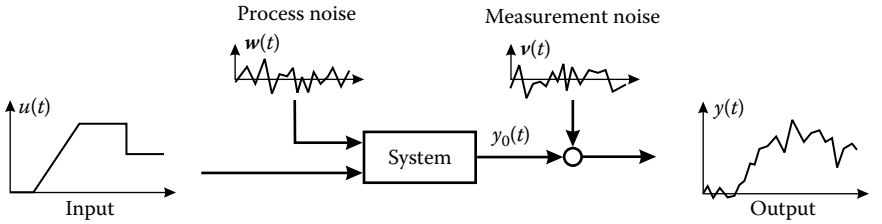


FIGURE 6.29 Kalman filter problem: the linear system has a deterministic known input $u(t)$, a stochastic unknown input $\boldsymbol{w}(t)$ (process noise), and measurement noise $\boldsymbol{\nu}(t)$ added to the system output $y_0(t)$.

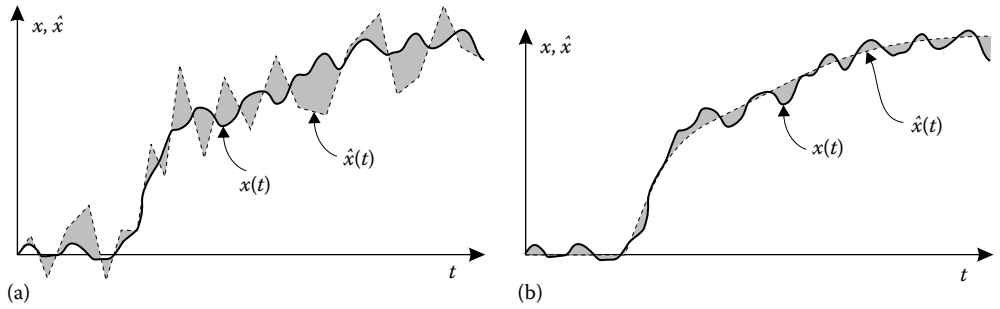


FIGURE 6.30 The problem of state estimation with measurement and process noise. (a) The measured output is a poor estimate. (b) The deterministic output is also a poor estimate.

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{v}. \tag{6.182}$$

Therefore, stochastic components of the measured output are partly actual system characteristic and partly noise. The objective is to design an observer such that the estimation error covariance is minimal. This task is not trivial, since taking the measurement as an estimate produces an estimation error equal to measurement noise, and taking the deterministic prediction as an estimate completely neglects the effects of process noise (Figure 6.30). Consequently, the optimal observer must consist of some form of compromise, which quantitatively incorporates the effects of both noise sources.

6.7.1.2 Analysis of Covariance

Since the process noise enters the system (6.181) as an input, the propagation and effect of its stochastic nature on the states must be described [1,5]. The goal is, therefore, to compute the covariance of the state vector in the presence of a stochastic system input $\boldsymbol{w}(t)$. The assumptions on the noise signals are uncorrelated zero mean ($\boldsymbol{\mu}_{\boldsymbol{w}(t)} = 0$, $\boldsymbol{\mu}_{\boldsymbol{v}(t)} = 0$) white noise sequences:

$$E\{\boldsymbol{w}(t)\boldsymbol{w}^T(\tau)\} = \boldsymbol{Q}_w\delta(t-\tau), \tag{6.183}$$

$$E\{\boldsymbol{v}(t)\boldsymbol{v}^T(\tau)\} = \boldsymbol{R}_v\delta(t-\tau), \tag{6.184}$$

$$E\{\boldsymbol{w}(t)\boldsymbol{v}^T(\tau)\} = \boldsymbol{R}_v\delta(t-\tau). \tag{6.185}$$

The definition for the covariance of the state vector x is

$$\boldsymbol{P}_x(t) = E\{\boldsymbol{x}(t)\boldsymbol{x}^T(t)\}. \tag{6.186}$$

A possible approach would be the direct computation of the integral; however, this would lead to a stochastic integration. Therefore, a differential equation for $\boldsymbol{P}_x(t)$ will be used instead. Direct differentiation of 6.186 gives

$$\begin{aligned}\dot{\boldsymbol{P}}_x &= E\left\{\dot{\boldsymbol{x}}\boldsymbol{x}^T + \boldsymbol{x}\dot{\boldsymbol{x}}^T\right\}\\ &= E\left\{\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^T + \boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{w}\boldsymbol{x}^T + \boldsymbol{x}\boldsymbol{w}^T\boldsymbol{G}^T\right\}\\ &= \boldsymbol{A}\boldsymbol{P}_x + \boldsymbol{P}_x\boldsymbol{A}^T + \boldsymbol{S} + \boldsymbol{S}^T,\end{aligned} \tag{6.187}$$

where the matrix $\boldsymbol{S}$ can be computed as

$$\begin{aligned}\boldsymbol{S} &= \boldsymbol{G}\int_{t_0}^{t} E\{\boldsymbol{w}(t)\boldsymbol{w}^T(t)\}\boldsymbol{G}^T\Phi^T(t-\tau)d\tau \\ &= \frac{1}{2}\boldsymbol{G}\boldsymbol{Q}_w\boldsymbol{G}^T\end{aligned} \tag{6.188}$$

due to the Dirac-delta function in 6.183. Hence, 6.187 can be written as

$$\dot{\boldsymbol{P}}_x = \boldsymbol{A}\boldsymbol{P}_x + \boldsymbol{P}_x\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{Q}_w\boldsymbol{G}^T; \qquad \boldsymbol{P}_x(t_0) = \boldsymbol{P}_{x0}. \tag{6.189}$$

This deterministic ordinary linear differential equation is the so-called Gauss–Markov equation. For a given initial condition $\boldsymbol{P}_{x0}$, the covariance matrix of the system states $P_x(t)$ can be computed. The stationary solution $\overline{\boldsymbol{P}_x}$ of 6.187 is given by

$$\boldsymbol{A}\overline{\boldsymbol{P}_x} + \overline{\boldsymbol{P}_x}\boldsymbol{A}^T = -\boldsymbol{G}\boldsymbol{Q}_w\boldsymbol{G}^T, \tag{6.190}$$

the so-called Lyapunov equation.

6.7.1.3 Derivation of the Filter Equations

The basic structure of the Kalman filter is identical to a Luenberger observer as introduced in 6.60:

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{H}(\boldsymbol{y} - \boldsymbol{C}\hat{\boldsymbol{x}}), \tag{6.191}$$

where the state estimation error is given by

$$\dot{\tilde{\boldsymbol{x}}} = (\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C})\tilde{\boldsymbol{x}} + \boldsymbol{G}\boldsymbol{w} - \boldsymbol{H}\boldsymbol{v}. \tag{6.192}$$

In both formulations (6.191) and (6.192), the so-called Kalman gain matrix H is the only unknown. The Kalman gain matrix H should now be chosen such that the error covariance matrix

$$\boldsymbol{P}(t) = E\{\tilde{\boldsymbol{x}}(t)\tilde{\boldsymbol{x}}^T(t)\} \tag{6.193}$$

will be minimal. Using 6.183, the Gauss–Markov equation for the error covariance is given by

$$\dot{\boldsymbol{P}} = (\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C})\boldsymbol{P} + \boldsymbol{P}(\boldsymbol{A} - \boldsymbol{H}\boldsymbol{C})^T + \boldsymbol{G}\boldsymbol{Q}_w\boldsymbol{G}^T + \boldsymbol{H}\boldsymbol{R}_v\boldsymbol{H}^T. \tag{6.194}$$

The minimal error covariance can be found by minimizing the trace of $\dot{\boldsymbol{P}}$ with respect to $\boldsymbol{H}$:

$$\frac{\partial Tr(\dot{\boldsymbol{P}})}{\partial \boldsymbol{H}} = -2\boldsymbol{P}\boldsymbol{C}^T + 2\boldsymbol{H}\boldsymbol{R}_v = 0. \tag{6.195}$$

The resulting optimal Kalman gain $\boldsymbol{H}$ is then given by

$$\boldsymbol{H} = \boldsymbol{P}\boldsymbol{C}^T\boldsymbol{R}_v^{-1} \tag{6.196}$$

and 6.194 can be written as

$$\dot{\boldsymbol{P}} = \boldsymbol{A}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{Q}_w\boldsymbol{G}^T - \boldsymbol{H}\boldsymbol{R}_v\boldsymbol{H}^T. \tag{6.197}$$

Equations 6.196 and 6.197 constitute the continuous-time Kalman filter. Together, they represent a Riccati equation structurally identical to 6.78, which was the solution for optimal control. This is not surprising since the Kalman filter is also the solution to a quadratic minimization problem with a linear side condition.

6.7.1.4 Properties of the Kalman Filter

The conditions for the existence and optimality of the Kalman filter are [1,5]

- Measurement and process noise are white stochastic processes.
- $\boldsymbol{R}_v$ must be positive definite (inversion in 6.196).
- $(\boldsymbol{A}, \boldsymbol{C})$ must be observable.
- $(\boldsymbol{A}, \boldsymbol{G})$ must be controllable.

If these preconditions are met, the Kalman filter has the following properties:

- The estimate is unbiased:

$$E\{\hat{\boldsymbol{x}}\} \rightarrow E\{\boldsymbol{x}\}. \tag{6.198}$$

- The estimate is optimal with respect to minimal error covariance:

$$E\{\tilde{\boldsymbol{x}}^T(t)\tilde{\boldsymbol{x}}(t)\} = min. \tag{6.199}$$

Therefore, each $E\{\tilde{x}_i(t)^2\}$, $i = 1, \ldots, n$ is minimal.

- The error covariance can be computed off-line using 6.197, without knowledge of actual measurements.
- The solution $\boldsymbol{P}(t)$ constitutes a self-diagnosis for the accuracy of the filter.

The Kalman gain is in general a time-dependent quantity. It should be noted that a stationary solution for the covariance matrix $\bar{\boldsymbol{P}}$ and the Kalman gain $\bar{\boldsymbol{H}}$ can be computed and a stationary Kalman filter can be implemented. In many practical problems, the stationary Kalman filter will be sufficient.

It can be shown [1,13] that the Kalman filter can also be applied to parameter estimation problems. Under certain conditions, it is equivalent to the RLS estimator but it can be used for combined state and parameter estimation. In this case, the original state vector x is augmented with the unknown parameters, the original system matrix $\boldsymbol{A}$ is augmented with an identity matrix block, and the output matrix $\boldsymbol{C}$ is left unaltered. Of course, observability and controllability of the augmented system must exist.

Example 6.18

A second-order continuous-time linear state-space system is given by

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -1 & 0.2 \\ 0.1 & -1.5 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\boldsymbol{u} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{w}, \quad \boldsymbol{y} = [0 \quad 1]\boldsymbol{x} + \boldsymbol{v}$$

and the covariance matrices are

$$\boldsymbol{Q} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.0001 \end{bmatrix}, \quad \boldsymbol{R} = 1 \times 10^{-5},$$

respectively. A stationary Kalman filter is to be designed and compared with two Luenberger observers.

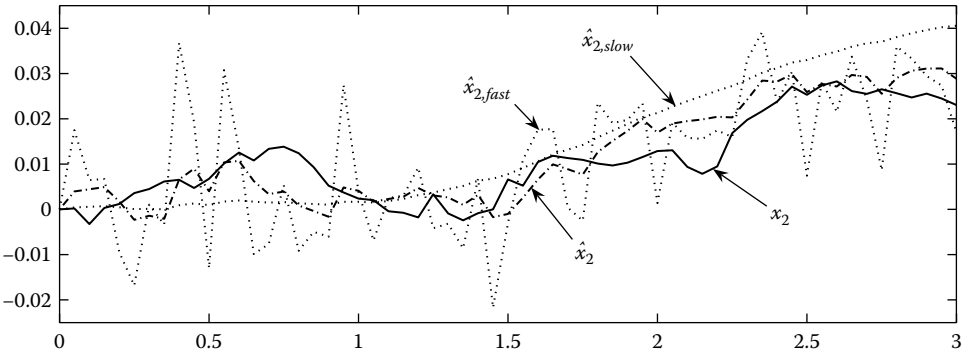


FIGURE 6.31 True state x_2 and estimations from Kalman filter ($\tilde{x}_2$), fast ($\tilde{x}_{2,fast}$), and slow ($\tilde{x}_{2,slow}$) Luenberger observers.

The stationary solution of the Riccati equation (6.197) yields the constant Kalman gain (6.196):

$$\boldsymbol{H} = [117.0 \quad 4.47]^T.$$

In a numerical simulation, the Kalman filter is compared to two Luenberger observers: one with a tenfold larger feedback gain (subscript "fast") and one with only one-tenth of the Kalman gain (subscript "slow"). The results are plotted in Figure 6.31. From the figure, it is clear that the fast observer is strongly compromised by the measurement noise while the slow observer is basically a deterministic prediction that mostly ignores the process noise. The optimality of the Kalman filter is reflected in the estimation error variances:

$$\sigma^2_{e_2} = 0.41.10^{-4}, \quad \sigma^2_{e_2,fast} = 1.41.10^{-4}, \quad \sigma^2_{e_2,slow} = 1.54.10^{-4}.$$

6.7.1.5 Stochastic Separation Principle

A stochastic separation is valid in analogy to the deterministic separation principle (see Equation 6.67). If the Kalman filter is combined, for example, with an optimal state-space controller, the resulting closed-loop system will minimize a criterion similar to 6.76, where the expectation operator has to be used additionally in order to incorporate the stochastic nature of the state vector. This result implies that for any given linear controller (which was designed using the deterministic assumptions of Section 6.3.3), the Kalman filter constitutes the observer which guarantees a minimal criterion (6.76) under the influence of noise.

6.7.1.6 Discrete-Time Kalman Filter

The discrete-time Kalman filter is an important formulation for real-time implementations. Based on the discrete-time representation of a state-space system

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{G}\boldsymbol{w}(k); \quad \boldsymbol{x}(k=0) = \boldsymbol{x}_0, \tag{6.200}$$

$$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{v}(k) \tag{6.201}$$

the recursive equations for the Kalman filter are

$$\boldsymbol{x}^*(k+1) = \boldsymbol{A}\hat{\boldsymbol{x}}(k) + \boldsymbol{B}\boldsymbol{u}(k), \tag{6.202}$$

$$\boldsymbol{P}^*(k+1) = \boldsymbol{A}\boldsymbol{P}(k)\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{Q}(k)\boldsymbol{G}^T, \tag{6.203}$$

$$\boldsymbol{K}(k+1) = \boldsymbol{P}^*(k+1)\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}^*(k+1)\boldsymbol{C}^T + \boldsymbol{R}(k+1)]^{-1}, \tag{6.204}$$

$$\hat{\boldsymbol{x}}(k+1) = \boldsymbol{x}^*(k+1) + \boldsymbol{K}(k+1)[\boldsymbol{y}(k+1) - \boldsymbol{C}\boldsymbol{x}^*(k+1)], \tag{6.205}$$

$$\boldsymbol{P}(k+1) = [\boldsymbol{I} - \boldsymbol{K}(k+1)\boldsymbol{C}]\boldsymbol{P}^*(k+1). \tag{6.206}$$

$\boldsymbol{P}^*(k + 1)$ and $\boldsymbol{x}^*(k + 1)$ are the a priori predictions of the error covariance and the state vector at sampling instant k, respectively [5].

6.7.1.7 Stability Problems and Divergence

The Kalman filter can be impaired by many problems: the process noise may not be white; measurement noise may be correlated; controllablility of $(\boldsymbol{A}, \boldsymbol{G})$ may be given but poor; and the process model may be inaccurate. In these cases, the Kalman filter may not converge or the error covariance estimate is biased. Remedies do exist [1,5], but in most cases a suboptimal filter is the result.

6.7.1.8 Related Filters for Nonlinear Systems

For nonlinear systems, the extended Kalman filter (EKF) can be applied to both state and parameter estimation. For a nonlinear system representation,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), t) + \boldsymbol{G}\boldsymbol{w}(t),$$

$$\boldsymbol{y}(t) = \boldsymbol{g}(\boldsymbol{x}(t)) + \boldsymbol{v}(t). \tag{6.207}$$

a local (state-dependent) approximation of the system is given by the Jacobians

$$\boldsymbol{A}(\hat{\boldsymbol{x}}, t) = \frac{\partial \boldsymbol{f}(\boldsymbol{x}(t), (t)}{\partial \boldsymbol{x}(t)}\Big\|_{\boldsymbol{x}(t)=\hat{\boldsymbol{x}}(t)}, \tag{6.208}$$

$$\boldsymbol{C}(\hat{\boldsymbol{x}}, (t)) = \frac{\partial \boldsymbol{g}(\boldsymbol{x}(t))}{\partial \boldsymbol{x}(t)}\Big\|_{\boldsymbol{x}(t)=\hat{\boldsymbol{x}}(t)},$$

which are evaluated at the current state estimate $\hat{\boldsymbol{x}}$. Using the Jacobians (6.209), the same equations as given for the ordinary Kalman filter are utilized [5,6].

Obviously, the EKF makes use of a local first-order approximation of the nonlinear functions $\boldsymbol{f}$ and $\boldsymbol{g}$. However, this may introduce large errors and due to the local linearization the Kalman gain itself is now a stochastic variable. The performance of the EKF is therefore strongly dependent on the type of the nonlinearities, and modeling errors easily lead to convergence problems (note that the observability of $(A(\hat{\boldsymbol{x}}, t), C(\hat{\boldsymbol{x}}, (t)))$ and the controllability of $(A(\hat{\boldsymbol{x}}, t), G)$, is now state dependent).

To overcome these problems, more advanced techniques have been proposed, like the unscented Kalman filter (UKF) [10] or the DD2 filter [17], where without the necessity of partial derivatives a more accurate and robust higher-order approximation is achieved. Another approach is the SMO, which is introduced in Section 6.7.2.

6.7.2 Sliding Mode Observer

6.7.2.1 Motivation

The structure and parameters of the model necessary for an observer are almost always uncertain. This may be due to unknown parameter values, nonlinearities, or time-varying systems. However, the Luenberger observer and the Kalman filter critically rely on a good prediction model. It is therefore desirable to have an estimation tool, which delivers unbiased state or parameter estimates even in the presence of unknown but bounded parameter uncertainties.

6.7.2.2 Principle of the SMO

The design of a SMO is structured in two parts:

1. A so-called sliding surface (a manifold whose dimension is less than that of the original system) is designed in order to assure the desired system properties (dynamics, stability).
2. A switched feedback signal is designed which steers the states (i.e., the estimation errors) to the manifold and keeps all subsequent motion on it.

6.7.2.3 Structure of a SMO

Observers based on the sliding mode concept can have a variety of structures [4,20]. Here, a Lyapunov-based observer for uncertain systems is presented. Consider a system given by

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t), \tag{6.209}$$

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}, \tag{6.210}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are constant matrices assumed to be known, and the unknown but bounded function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)$ represents nonlinearities, time-varying terms, and disturbances (e.g., unknown inputs, uncertain parameters). The pair $(\boldsymbol{A}, \boldsymbol{C})$ is assumed to be observable.

Defining the output error $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ with $\hat{\boldsymbol{y}} = \boldsymbol{C}\hat{\boldsymbol{x}}$, a SMO for the earlier defined system is given by

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{H}\boldsymbol{e} + \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}, t), \tag{6.211}$$

where the switching function $\boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}, t)$ is defined as

$$\boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}, t) = \rho \boldsymbol{P}^{-1}\boldsymbol{C}^T \frac{\boldsymbol{e}}{\|\boldsymbol{e}\|} \tag{6.212}$$

and in the case of a scalar output y and output error e:

$$\boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}, t) = \rho \boldsymbol{P}^{-1}\boldsymbol{C}^T sign(e). \tag{6.213}$$

The gain matrix $\boldsymbol{H}$ and an arbitrary symmetric positive definite matrix $\boldsymbol{Q}$ must be chosen so that the following conditions are satisfied [4]: $\boldsymbol{A}_0 = \boldsymbol{A} - \boldsymbol{H}\boldsymbol{C}$ must be stable and $PA_0 + A_0^T P = -Q$. If ρ is chosen according to

$$\rho > \left\|\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}, t)\right\|, \quad \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) = \boldsymbol{P}^{-1}\boldsymbol{C}^T\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}, t), \tag{6.214}$$

quadratic stability of the observer is guaranteed. In practical problems, the value of ρ is often chosen as a compromise between robustness and noise attenuation.

An ideal sliding mode requires infinite switching frequency. Whenever the switching frequency is limited "chattering" occurs, that is, the switched feedback signal is visible in the state estimates.

Example 6.19

A second-order continuous-time nonlinear state-space system is given by

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -1 & 0.2 \\ 0.1 & -0.3 \end{bmatrix}\boldsymbol{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u + \begin{bmatrix} 0 \\ -1.2atan(x_2) \end{bmatrix}, \quad y = [0 \quad 1]\boldsymbol{x}.$$

Linearization around $\boldsymbol{x} = [0\ 0]^T$ yields the deterministic part of the system given in Example 6.18. An SMO is to be designed and compared with a Luenberger observer (LO) for a nonstationary input.

The feedback gain matrix $\boldsymbol{H}$ and the symmetric positive definite matrix $\boldsymbol{Q}$ from Equation 6.211 are chosen as

$$\boldsymbol{H} = [117.0 \quad 4.47]^T, \quad \boldsymbol{Q} = \begin{bmatrix} 120 & 3.45 \\ -0.03 & -0.182 \end{bmatrix}$$

and the switched signal according to Equation 6.213 is given by

$$\boldsymbol{S}(\boldsymbol{x}) = [57.2 \quad 10.3]^T \, sign(e).$$

In a numerical simulation, the SMO is compared to the LO with a feedback gain of $\boldsymbol{H} = [117.0 \quad 4.47]^T$. The results are plotted in Figure 6.32. The LO produces a biased estimate in both state variables, while the SMO is capable of reconstructing $\hat{x}_1$ within a small error bound and $\hat{x}_2$ with an error only due to chattering.

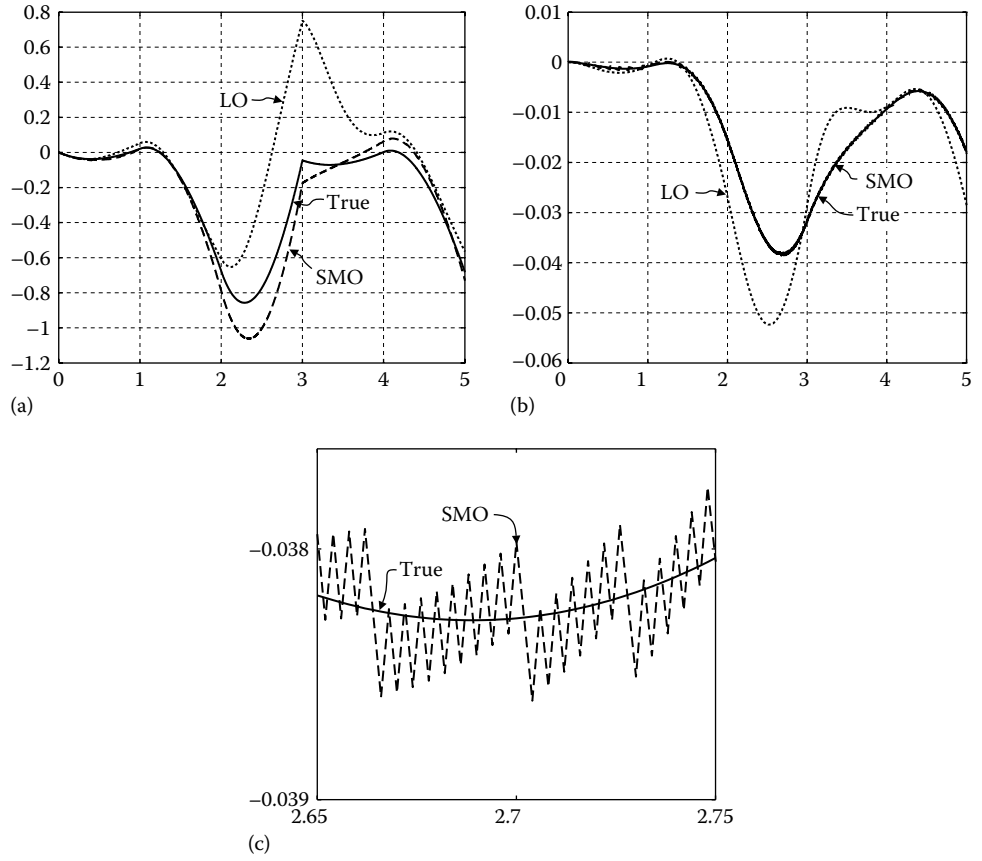


FIGURE 6.32 (a) True state x_1 and estimations from SMO and LO. (b) True state x_2 and estimations from SMO and LO. (c) Chattering in the SMO estimate $\hat{x}_2$ due to finite switching frequency.

REFERENCES

1. G. Arthur. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 2001.
2. J.S. Bendat and A.G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, New York, 1980.
3. J.S. Bendat and A.G. Piersol. *Random Data*. John Wiley & Sons, New York, 1986.
4. Ch. Edwards and S.K. Spurgeon. *Sliding Mode Control: Theory and Applications*. Taylor & Francis, London, U.K. 1998.
5. M.S. Grewal and A.P. Andrews. *Kalman Filtering: Theory and Practice using MATLAB*. John Wiley & Sons, New York, 2001.
6. S. Haykin. *Adaptive Filter Theory*. Prentice Hall, Upper Saddle River, NJ, 1996.
7. S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Englewood Cliffs, NJ, 1999.
8. A. Isidori. *Nonlinear Control Systems 1. An Introduction*, 3rd edn. Springer, New York, 1995.
9. J.-J.E. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, Englewood Cliffs, NJ, 1991.
10. S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, FL, 1997.
11. K. Hassan. *Nonlinear Systems*. Prentice Hall, Englewood Cliffs, NJ, 2001.
12. J. Kautsky, N.K. Nichols, and P. Van Dooren. Robust pole assignment in linear state feedback. *International Journal of Control*, 41(5):1129–1155, 1985.
13. L. Ljung. *System Identification: Theory for the User*, 2nd edn. Prentice Hall, Englewood Cliffs, NJ, 1997.
14. L. Ljung and T. Glad. *Modeling of Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1994.
15. M. Moonen and J. Ramos. A subspace algorithm for balanced state space system identification. *IEEE Transactions on Automatic Control*, 38(11):1727–1729, November 1993.
16. O. Nelles. *Nonlinear System Identification*, 1st edn. Springer, Berlin, Germany, 2002.
17. M. Norgaard, N.K. Poulsen, and O. Ravn. New developments in state estimation for nonlinear systems. *Automatica*, 36(11):1627–1638, 2000.
18. A.V. Oppenheim and R.W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
19. J.G. Proakis and D.G. Manolakis. *Digital Signal Processing*. Prentice Hall, Upper Saddle River, NJ, 1996.
20. V. Utkin, J. Guldner, and J. Shi. *Sliding Mode Control in Electromechanical Systems*. CRC Press, Boca Raton, FL, 1999.
21. K. Zhou, J.C. Doyle, K. Glover, and J.C. Doyle. *Robust and Optimal Control*, 1st edn. Prentice Hall, Englewood Cliffs, NJ, 1995.

7 Actuators and Sensors

Yoshihiro Suda

CONTENTS

7.1 Introduction 221
7.2 Vehicle Dynamics and Control 222
7.2.1 Classification of Controls 222
7.2.2 Suspension Control 222
7.2.3 Steering Control 223
7.2.4 Drive Control 223
7.2.5 Sensors and Actuators in Vehicle Dynamics Controls 224
7.3 Modeling of Actuators 225
7.3.1 Classification of Electric Motors 225
7.3.2 DC Motors 225
7.3.3 AC Motors 227
7.3.4 Stepping Motors 227
7.3.5 Conversion of Rotational to Transversal Motion 228
7.3.6 Hydraulic Actuators 230
7.4 Applications of Actuators 230
7.4.1 Vibration Control 230
7.4.2 Steering Control 231
7.4.3 Active Front Steering Angle Control System 233
7.5 Sensors for Automobile Control 235
7.5.1 Sensing 235
7.5.2 Sensors for Detecting Own Vehicle Condition 236
7.5.2.1 Acceleration Sensor 236
7.5.2.2 Yaw-Rate Sensor 237
7.5.2.3 Force and Torque Sensor 237
7.5.3 Sensors for Detecting Vehicle Environment 237
7.5.3.1 Millimeter-Wave Radar 237
7.5.3.2 Laser Radar 238
7.5.3.3 Sonar 238
7.5.3.4 Vision Sensor 238
Acknowledgments 238
References 238

7.1 INTRODUCTION

Actuators and sensors are important devices for vehicle dynamic control (VDC). In the usual feedback control system, it is important to detect the controlled values, such as vehicle motions, velocities, and yaw rate. The output of control system should be applied the force or torque in the dynamic system. In the vehicle system, dynamic experiments are valuable methods for research and development. In that time, the vehicle motion must be detected without any control system. Sensor system is necessary in that case.

For the design of controller and modeling of vehicle, the simple model is used as the first step of examination. In that case, we assume that the controlled value is ideally measured and the output of controller also ideally transmits to the dynamic system. However, the actuator and sensors are dynamic system themselves. They have dynamic properties. It is important to consider such dynamic properties in the detailed examination of the VDC and measurements.

In this chapter, at first, the vehicle dynamics and control are explained to understand the actuators and sensors. Next, the modeling of actuator is shown. Some applications of actuators for vehicle system are also shown. For sensors, some remarks of sensing system are decried. Finally, some applications of sensors in vehicle control, vehicle measurements, and vehicle environment measurements are shown.

7.2 VEHICLE DYNAMICS AND CONTROL

7.2.1 Classification of Controls

In the design of automobiles, various electric devices have been developed for realizing further safety, comfort, and availability of vehicle. Control systems composed with these devices can be roughly categorized as the power train control, the electro-communication, the vehicle body control, and the chassis control. The chassis control is also called as vehicle dynamics control. In this chapter, the control is treated for surveying recent actuator and sensor system technologies (Figure 7.1).

VDCs fundamentally aim to control the automobile basic motion (i.e., running, curving, and braking), and improves further handling stability, riding comfort, and safety. The control of the vehicle dynamics can also be classified in detail as the suspension control, the steering control, and the drive control. The suspension control, which control vertical, rolling, and pitching motion of vehicle, aims to improve riding quality and wheel tracking ability along road. The steering control aims to improve handling and stability on curving and controls yawing and lateral motion of vehicle. Drive control aims vehicle longitudinal motion control.

7.2.2 Suspension Control

Automobile suspension has two functions: one is absorbing vibration for improving riding comfort and the other is transmitting traction and braking force aroused between wheel and road for improving handling stability. From the view point of riding comfort, soft spring, which isolate severe stimulus from road, is necessary. On the other hand, hard suspension is necessary for realizing excellent handling stability. The compatibility for both requirements is quite crucial problem to be solved in the design of suspension.

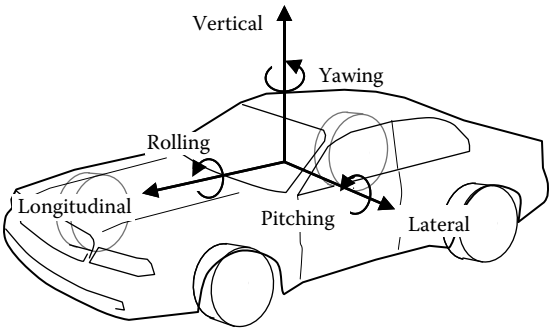


FIGURE 7.1 Controlled degree of freedom of automobile with vehicle dynamics control systems.

The definition of the active suspension is shown from the points of the power source and control function according to the road condition. The semi-active suspension is also defined as the suspension which does not have power source but has control function according to the road condition. These controlled suspensions are realized by utilizing hydraulic and pneumatic actuators. Control method of suspension has started with skyhook control which is useful for both active and semi-active suspensions control algorithm. In the control of suspensions, many researches have been reported such as optimal control, robust control, and adaptive control.

Even though the hydraulic actuators had been mainly used in the suspension control of the vehicle, the electromagnetic suspension has also been developed. The electromagnetic suspension gives many benefits in simplicity for damping tuning, maintenance, and high precision control and response. Furthermore, the possibility of self-powered control, which utilizes vibration energy regenerated by suspension and stored in capacitor for using active control, has been studied.

7.2.3 Steering Control

The control of four-wheel steering (4WS) [1] had been developed from the early 1970s. In the first step of the control algorithm for rear steering angles, simple feedback controls regarding such as front steering angle, steering force, and yaw-late were considered. As the relation between control theory and vehicle dynamics had been gradually clarified, more effective feedback control algorithm was developed. For example, the control where the output signal with sideslip angle of center of gravity is used for output zeroing was considered. Model following control, where the error signal between reference model and the output signal composed with vehicle lateral acceleration and yaw-late sensors, was also proposed. Previously mentioned 4WS controls are distinct from what is chosen as controlled output, and lead great difference for stability factor, handling stability, and disturbance response. The more control algorithm becomes complicated, the more sensors are required to be equipped. The work of 4WS had theoretically summarized by the end of 1980, and leads vehicle stability control (VSC) and direct yaw-moment control (DYC) [2–5].

In recent years, the application of fly-by-wire (FBW) in aerospace category has been positively introduced into the steering control of automobiles. The steer-by-wire (SBW) [6,7], which is capable of flexible change of the relation between vehicle and driver, gives great possibility on vehicle handling control such as variable gear ratio control.

7.2.4 Drive Control

The basic function of the drive control is longitudinal control such as braking, acceleration, and keeping constant velocity. These controls are realized by anti-lock brake system (ABS), traction control system (TCS), and adaptive cruise control (ACC) [8,9], respectively. Initially, the ABS control mainly aimed to keep handling stability during full braking while wheels are locked. The state of locking of wheel is indirectly measured by comparing vehicle velocity and wheel rotational velocity. In the estimation of absolute vehicle velocity from wheel rotational velocity, the accuracy of the measurement is still problem to be solved. If the brake pressure induced by the driver is such that the wheels lock, then the brake pressure must be reduced to regain steer ability. For this task, ABS uses a hydraulic unit, which has electromagnetic valves to keep the pressure in the wheel brakes.

If the driven wheels spin due to excess engine drive torque, handling becomes difficult. This has led then to the introduction of TCS which also preserve a high level of handling performance during driving with excess engine drive torque. The main task of the control algorithm is to keep a high level of traction force while at the same time keeping a sufficient level of lateral force generation by

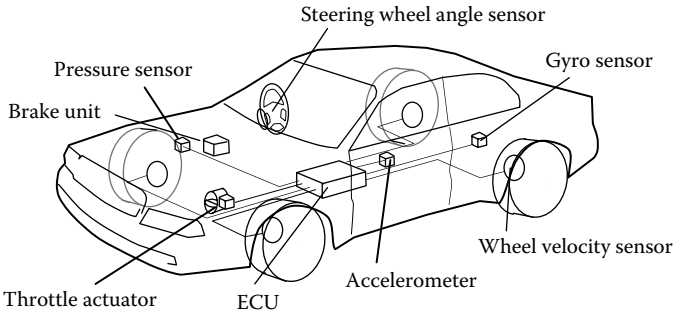


FIGURE 7.2 An example of sensor and actuator systems on ESP.

steering to preserve a high level of handling performance. Fortunately, for single axle-driven cars, the speeds of the non-driven wheels may be used to compute the free rolling speed of the driven wheels. Therefore, slip control of the driven wheels becomes possible.

Crouse control system adjusts the engine drive torque using throttle valve to keep the vehicle velocity to the desired value. As an example of vehicle velocity sensors used here, magnetic rotary encoder mounted on transmission is used.

The control of yaw momentum of vehicle is realized by controlling each wheel slip with longitudinal tire forces. This control leads stable vehicle yawing and lateral motion and improves steer ability in large region of slipping. The control is called "Electric Stability Program" (ESP; sometimes called as VDC, VSC, and so on, depending on car manufactures) and nowadays widely implemented on automobiles. Feedback control of the vehicle motion is possible by extending the TCS and ABS with additional sensors. As shown in Figure 7.2, ESP consists of many sensors as yaw-late sensors, accelerometer, steering wheel angle sensor, brake pressure sensor, and tachometer. Steering wheel angle sensor detect the curving trajectory on which driver desires to trace, then comparing the error between measured vehicle trajectory using yaw-late sensor and accelerometer, and controls the sideslip with feedback control.

Direct Yaw Control (DYC) also gives yaw-late and lateral acceleration control with longitudinal tire force. The purpose of DYC is almost same as ESP, where the difference of traction of each wheel aroused by electrically controlled Limited Slip Differential (LSD) or other differential torque devices allows more sufficient force generation than that of ESP. In recent years, electric vehicle with in-wheel drive motor is considered. The vehicle is essentially capable of high performance of DYC with motor torque control.

As in the case of SBW, throttle-by-wire (TBW) is also developed and allows cooperative use with stability controls (ESP, DYC, and so on). The cooperative use of drive control and steering control is also considered for realizing further high performance of steer ability of vehicle. Since both of control realize yaw-late and lateral acceleration control, the dynamics coupling, which results from connection of each interactive subsystem, becomes problem to be solved. In order to solve the problem, integrated vehicle control is proposed and is the latest topic of the vehicle dynamics control [10,11].

7.2.5 Sensors and Actuators in Vehicle Dynamics Controls

Some example of sensors used in previously mentioned controls are steering sensor, vehicle velocity sensor, tachometer, accelerometer, yaw-late sensors, and so on. Some of these sensors are used in more than two controls. These devices are selected along with control performance such as time-response and frequency characteristics. Furthermore, vehicles have special actuators such as ABS, LSD engine drive torque, and so on. Table 7.1 gives the classified according to controls, controlled variables, actuators, and sensors.

TABLE 7.1
Scope of Control System and Corresponding Controlled Variable

	Control	Variable	Actuator	Sensor
Suspension control	Vibration control, height control, antiroll control, etc.	Vertical, rolling, pitching motion	Hydraulic, pneumatic, electromagnetic	Accelerometer, stroke sensor, etc.
Steering control	EPS, 4WS, SBW (AFS, ARS)	Lateral, yawing motion	Hydraulic, pneumatic, electric actuator	Yaw-rate sensor, steering wheel angle, accelerometer, tachometer, etc.
Drive control	ABS, TCS, adaptive cruise control, vehicle stability control, direct yaw control	Longitudinal, yawing motion	Engine drive torque, LSD, ABS, etc.	Yaw-rate sensor, tachometer, steering wheel angle, brake pressure, throttle sensor, etc.

7.3 MODELING OF ACTUATORS

7.3.1 Classification of Electric Motors

As mentioned earlier, several kinds of actuators are utilized in automobiles. They improve running performance, ride comfort, and operate many other equipments. Major actuators utilized in the automobiles are electric motors and hydraulic actuators. Electric motors are transducers that convert electric energy into mechanical energy. The electric motors are divided into three categories: direct current (DC), alternative current (AC), and stepping motors [12].

7.3.2 DC Motors

Torque of the DC motors originates from electromagnetic force acting on a current-carrying conductor in a developed magnetic field. A typical DC motor produces torque τ_r, when the electric current passes in its armature coil:

$$\tau_r = \phi_\tau i_r. \tag{7.1}$$

To the contrary, when the rotor rotates at the angular speed of ω_r, voltage e_b is induced in the coil as

$$e_b = -\phi_e \omega_r. \tag{7.2}$$

This voltage is called back electromotive force (emf). The constants ϕ_τ and ϕ_e have the same values, since consistency of mechanical and electrical work must be maintained. Then we symbol them by ϕ ($= \phi_\tau = \phi_e$) [13].

The magnetic field of a DC motor is produced by permanent magnets (PMs) or field windings. According to the methods to produce the magnetic field, DC motors are mainly classified to three categories: PM motors, series motors, and shunt motors. Since a PM motor develops the magnetic field with a PM, it can operate with high efficiency and has a linear characteristic, which makes it control-friendly. A PM motor is mainly utilized for a servomotor in automobiles [14]; therefore, only DC motors with PM are treated in this section.

Figure 7.3 shows a model of a geared PM-type DC motor.

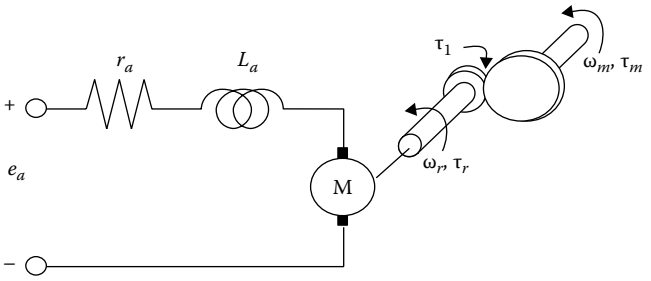


FIGURE 7.3 Model of a PM DC motor.

The parameters are defined as

i_a is the current of the armature
e_a is the applied voltage
e_b is the back emf
r_a is the resistance of the armature
L_a is the inductance of the armature
ω_r is the angular velocity of the rotor
τ_r is the torque resulted from electromagnetic force
τ_1 is the load torque of the rotor
τ_m is the load torque of the output shaft
ω_m is the angular velocity of the shaft
G is the gear ratio
J_r is the inertia of the rotor
J_m is the inertia of the output shaft

A circuit equation is

$$L\frac{di_a}{dt} + r_a i_a = e_a - e_b. \tag{7.3}$$

Moment equations of the rotor and the output shaft are

$$J_r \frac{d\omega_r}{dt} = \tau_r - \tau_1 \tag{7.4}$$

$$J_m \frac{d\omega_m}{dt} = G\tau_1 - \tau_m. \tag{7.5}$$

The relationship between τ_r and τ_m is

$$\omega_m = \frac{\omega_r}{G}. \tag{7.6}$$

Then the moment equations derived as

$$\left(J_m + G^2 J_r\right)\frac{d\omega_m}{dt} = G\tau_r - \tau_m. \tag{7.7}$$

From Equations 7.1 through 7.3, 7.6, and 7.7, the state equations are written in vector-matrix form:

$$\begin{bmatrix} \dfrac{di_a}{dt} \\ \dfrac{d\omega_m}{dt} \end{bmatrix} = \begin{bmatrix} -\dfrac{r_a}{L_a} & -\dfrac{G\phi}{L_a} \\ \dfrac{G\phi}{J_m + G^2 J_r} & -\dfrac{G^2\phi}{J_m + G^2 J_r} \end{bmatrix} \begin{bmatrix} i_a \\ \omega_m \end{bmatrix} + \begin{bmatrix} \dfrac{1}{L_a} \\ 0 \end{bmatrix} e_a + \begin{bmatrix} 0 \\ \dfrac{-1}{J_m + G^2 J_r} \end{bmatrix} \tau_m. \tag{7.8}$$

7.3.3 AC Motors

Majority of AC motors utilized as servomotors in automobile systems are synchronous motors whose magnetic field is produced by PMs [14]. The main difference between this type of AC motors and DC motors is that they do not have brushes, which commutate the armature current mechanically; therefore, they are also called brushless DC motors. Although the commutations need to be done electrically, they do not need a brush to wear out and this makes the motor maintenance-free. Figure 7.4 shows basic operational principles of the motor. In the brushless motor, the rotor consists of PMs and windings are located external to the rotating part. Hall-element sensors are placed near the rotor and measure the magnetic flux to estimate the rotor angle. According to the output of the Hall element, the circuit is changed by some kinds of transistors and the current is commutated. Although the brushless DC requires additional energy to operate sensors and transistors, it is maintenance-free and its rotor inertia is small compared with that of brushed motors. Recently, brushless DC motors have become popular electric motors and have been utilized for drive motors of a hybrid car as well as servomotors [15].

7.3.4 Stepping Motors

Stepping motors convert digital pulses into mechanical rotation. The amount of rotation is proportional to the number of pulses and the speed of rotation is relative to the frequency of those pulses. Since position error is limited and known, stepping motor has high reliability for position control.

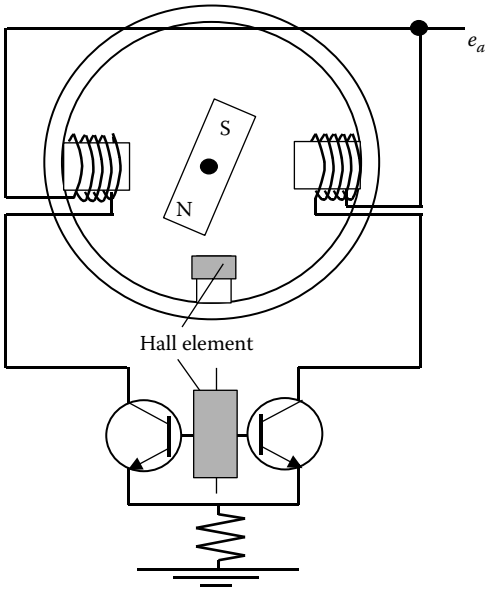


FIGURE 7.4 Basic operational principles of synchronous AC motors.

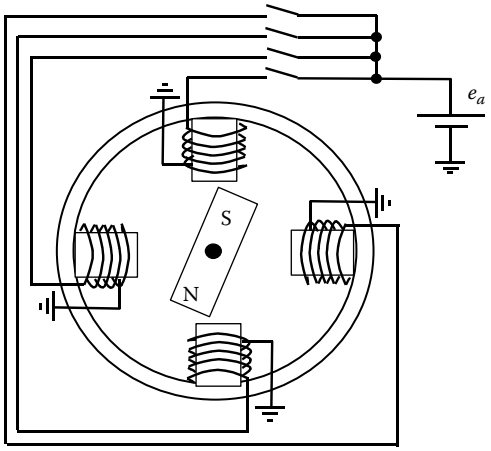


FIGURE 7.5 Basic operational principles of step motors.

In automobile system, they are utilized for meters on instrument panels, throttle valve control systems, and so on. Basic operational principles are shown in Figure 7.5. It consists of PM rotor and stator coils. Rotation is achieved by switching the voltage source to each winding in turn.

7.3.5 Conversion of Rotational to Transversal Motion

The majority of electric actuators utilized in automobiles are powered by rotary motors. Then, it is often necessary to convert rotational motion to transversal motion when we use them to produce the translational forces. One of the major converters is a ball-screw-and-nut [16]. Figure 7.6 shows one of the possible configurations of the electric motor assembled with a ball-screw-and-nut. By this mechanism, output torque of the rotary motor τ_m is converted to output force f as

$$f = \frac{2\pi}{l}\left(\tau_m - J_s \frac{d\omega_m}{dt}\right) \tag{7.9}$$

where J_s and l are inertia moment of the screw and lead of the screw. And when the speed of transversal motion is $\dot{z}$, angular speed of the screw ω_m is obtained by

$$\omega_m = \frac{2\pi}{l}\dot{z}. \tag{7.10}$$

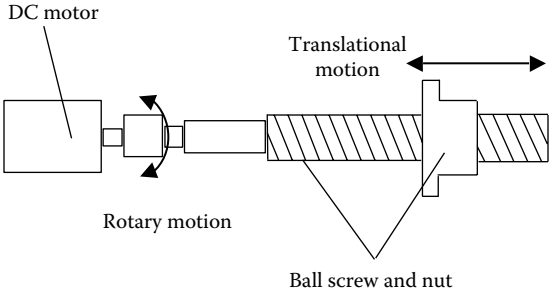


FIGURE 7.6 Configuration of an electric motor assembled with a ball-screw-and-nut.

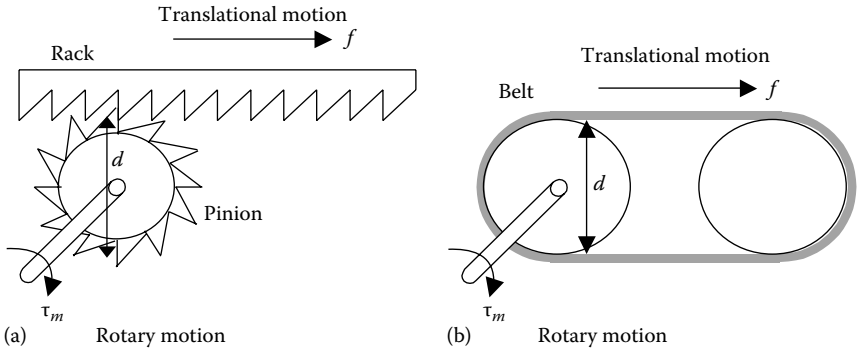


FIGURE 7.7 Conversion mechanisms: (a) a rack-and-pinion-gear and (b) a belt-and-pulley system.

Other conversion mechanisms are a rack-and-pinion-gear and belt-and-pulley system as shown in Figure 7.7. When the inertia moment and the diameter of a pinion or a pulley are J_p and d, respectively, the output force is obtained by

$$f = \frac{2}{d}\left(\tau_m - J_p \frac{d\omega_m}{dt}\right). \tag{7.11}$$

When the speed of transversal motion is $\dot{z}$, the angular speed of the motor shaft ω_m is derived as

$$\omega_m = \frac{2}{d}\dot{z}. \tag{7.12}$$

Using electric linear motors is also the possible way to produce translation forces. Figure 7.8 shows the configuration of a typical linear DC motor. This motor can produce the translational force proportional to the electric current in the motor coil without effects of friction and inertia forces of the conversion mechanism; therefore it is suitable for the actuator of the speaker, which needs high responsibility and preciseness. However, the power produced by the linear motor is small compared with the rotary motor of the same mass. The application of the linear motor in the automobile system is limited, although the innovation might change the situation.

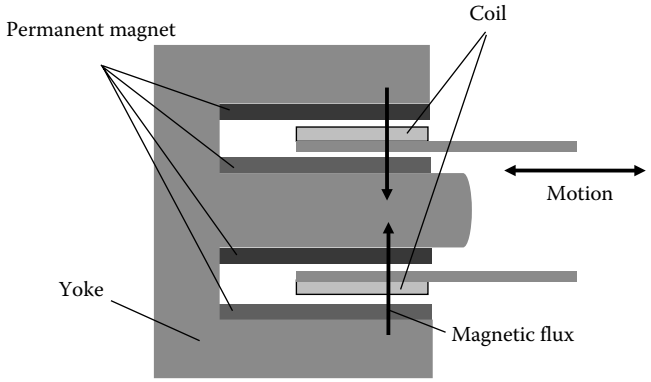


FIGURE 7.8 One of the configurations of linear DC motors.

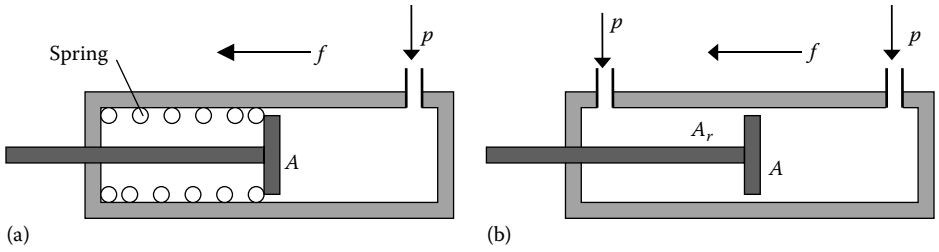


FIGURE 7.9 Configuration of a hydraulic actuator (a) is a single acting hydraulic actuator and (b) is a double acting hydraulic actuator.

7.3.6 Hydraulic Actuators

Hydraulic actuators are normally used when a large amount of force is required. Although there are several kinds of hydraulic actuators, such as a hydraulic cylinder, a rotary actuator, and a hydraulic motor, we concentrate on a hydraulic cylinder, which produces transversal motion and is frequently utilized as an actuator of an automotive system. Figure 7.9 shows configurations of the hydraulic cylinders. Figure 7.9a represents a single acting cylinder. Pressured oil is supplied to one side of the cylinder and output force f is obtained as

$$f = A\,p - f_{spring} \tag{7.13}$$

where p, A, and f_{spring} are pressure of the supplied oil, sectional area of the piston, and spring force. However, this type of the cylinder cannot produce force in the opposite direction. Usually, the piston is returned to its initial position by returning force produced by an internal spring. Figure 7.9b illustrates a double acting cylinder. This type of the cylinder has ports of the pressured oil in both sides of the cylinder; therefore, it can produce force in both directions. When the pressured oil is supplied to the left port, output force f is derived as

$$f = A\,p. \tag{7.14}$$

As the oil is supplied to the right port, the output becomes

$$f = -\left(A - A_r\right)p, \tag{7.15}$$

where A_r is sectional space of the rod.

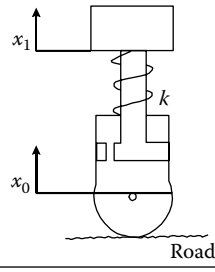
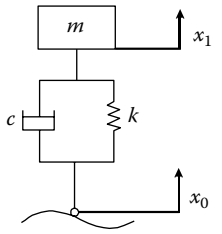
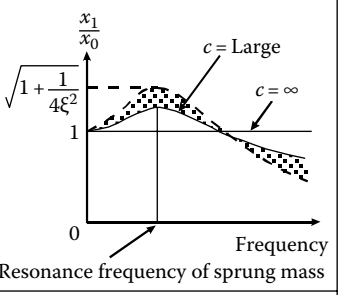
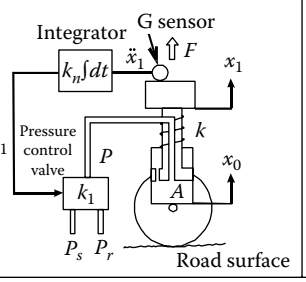
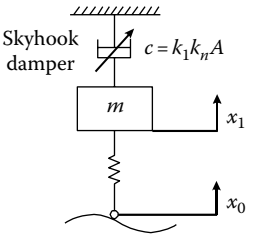
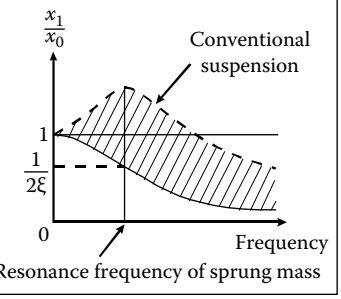
7.4 APPLICATIONS OF ACTUATORS

7.4.1 Vibration Control

In a conventional suspension of an automobile consisted of springs and dampers, it is a well-known fact that higher damping ratio of damper improves the vibration isolation around resonance frequency; however, it worsen the ride quality in high frequency range at the same time.

Active suspension gives a solution to this conflicting problem. Table 7.2 indicates the constitution and the performances of a conventional suspension and an active suspension, and it is shown that vibration isolation can be achieved both in resonance frequency and in high frequency range simultaneously by active suspension. The active suspension is considered to be effective for the vibration up to 30 Hz.

TABLE 7.2
Comparison between Conventional and Active Suspensions

	Structure	Equivalent model	Transfer characteristic
Conventional suspension			
Hydraulic active suspension			

As a control theory of the active suspension, skyhook damping theorem proposed by Karnopp et al. [17] is frequently used; in addition, other control theories such as optimal control theory, *H*-infinity control theory, and sliding mode theory are also applied in order to obtain higher performance.

In commercialized active suspension system, hydraulic actuator and pneumatic actuator are used. Figure 7.10 shows the layout of actuators and sensors for active suspension system with hydraulic system. Hydraulic system is consisted with an oil pump, a reserve tank, accumulators, controlled valve units, and actuators. As for sensing system, a certain number of three-axes accelerometers are located in both sprung mass (car body) and unsprung mass (knuckle, suspension arms, damper tube, etc.), and yaw-rate sensor, stroke sensor, and load cells are also used. Other information such as steering angle and velocity are also used for active control.

The active suspension system is also used for purposes other than vibration control. Attitude control is one of the applications. Roll and pitch motion of an automobile can be controlled by same active suspension units. In addition, vehicle height can be also controlled with same devices, so that the clearance to the ground can be generated. It is efficient especially in rough road.

Disadvantage of the active suspension system is energy consumption. It requires considerable amount of energy. Semi-active suspension is also developed from this point of view, which changes damping characteristics of damper, not generating vertical forces directly. Semi-active suspension also has a merit that control system is rather simple in comparison with active suspension.

7.4.2 Steering Control

Electric Power Steering (abbreviated as EPS, Figure 7.11) system with steering assistance power supplied from an electric motor has been recently developed and mass-produced instead of the conventional hydraulic power steering system, since it has advantages concerning environmental preservation, energy conservation, and expanded degree of freedom of steering system layout

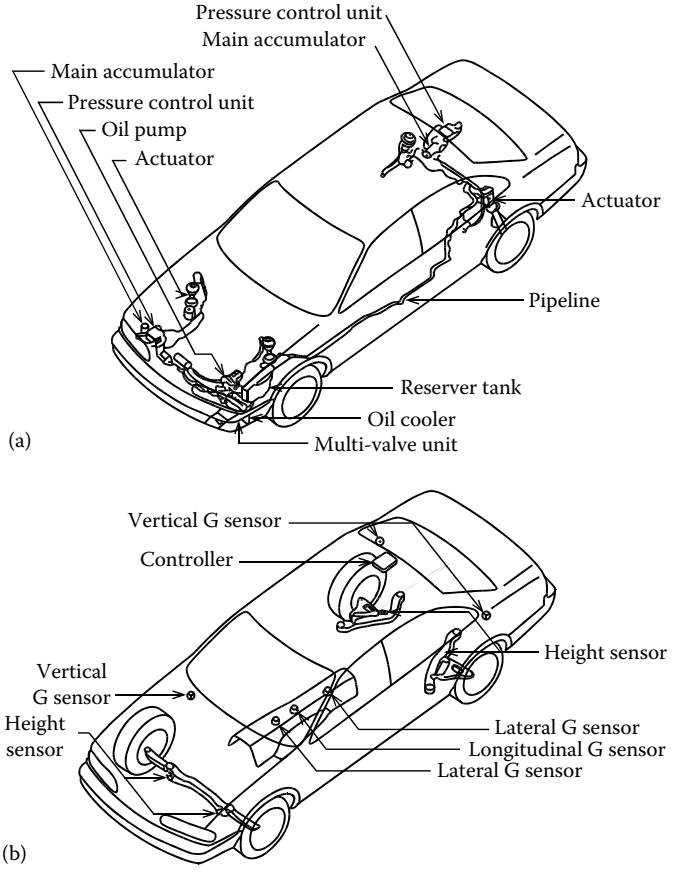


FIGURE 7.10 Layout of (a) actuators and (b) sensors.

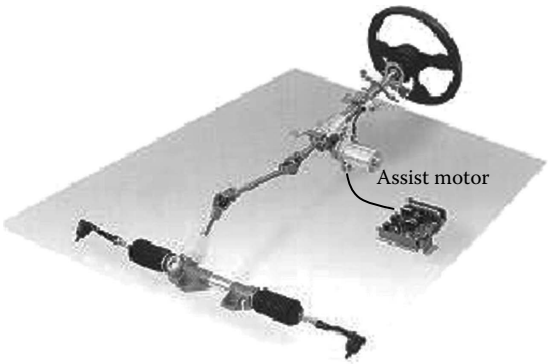


FIGURE 7.11 Column assist type EPS.

without hydraulic pump and pipes. In addition, the reduction of driver workload, and the grasping of vehicle situations and driving conditions by using EPS can contribute to the improvement of driving safety and the comfort in maneuvering [18]. However, a mechanical linkage existing between front tires and steering wheel makes the steering dynamics and vehicle dynamics unable to be designed independently. Therefore, to improve both the vehicle maneuverability and the steering feeling in maneuvering characteristics is very difficult.

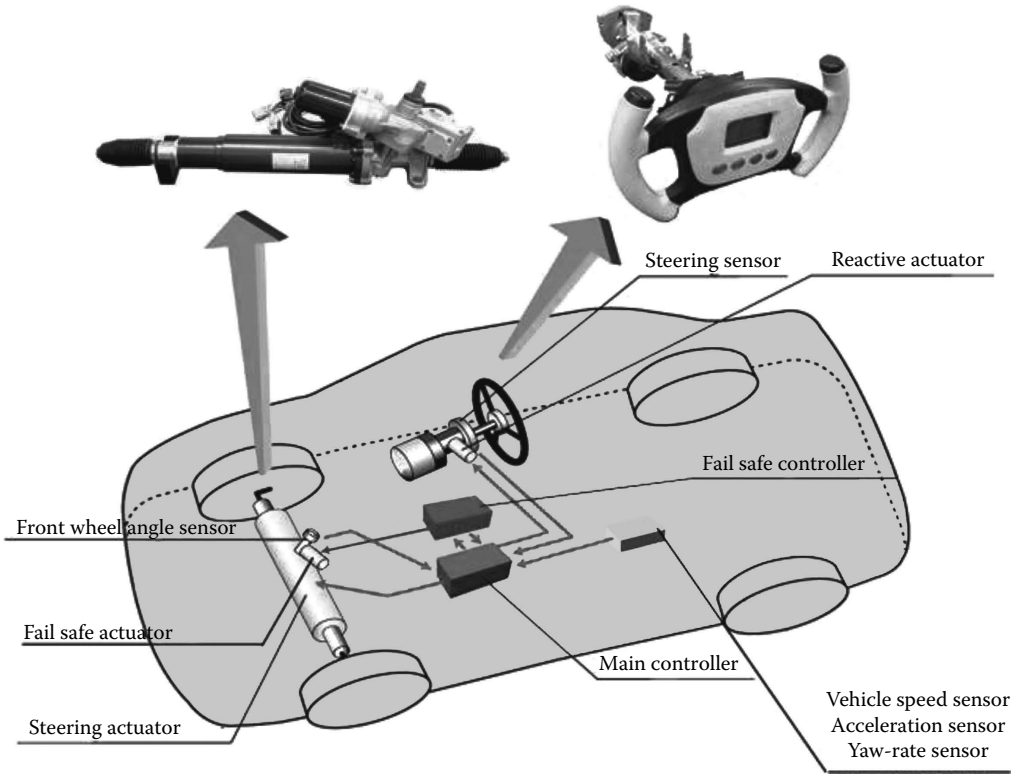


FIGURE 7.12 An example of steer-by-wire system. (From Segawa, M. et al., *JSAE Rev.*, 22, 383, 2001.)

To overcome this problem, SBW system becomes one of the solutions since the mechanical linkage does not exist as shown in Figure 7.12 [19]. Typically, SBW system has several merits as follows: (1) avoiding steering column shaft collision to driver, (2) flexibility of interior layout and steering situation, (3) flexibility of steering configuration design: for example, joystick, wheel, ellipse [19–21]. Moreover, SBW can control steering angle automatically to stabilize vehicle behavior and, can control steering characteristics such as decreasing the physical workload and transmitting the information of vehicle motion and road condition via a steering reaction torque back to the driver, without the mutual interference between the vehicle dynamics and the steering dynamics [22–24]. From the viewpoint of human–machine interface, SBW has structure merits in avoiding interference between the driver and steering wheel since the mechanical linkage between steering wheel and tire is replaced by electric wire system, consisting of sensors, steering actuators, and controllers [19,22].

As shown in Figure 7.12, the vehicle with SBW system consists of a reaction torque actuator for adjusting steering feeling and a front-tire steering actuator for steering the front tires.

7.4.3 Active Front Steering Angle Control System

For enhancing handling and stability of vehicle, the yaw rate and the lateral acceleration of vehicle are controlled to trace the desired value by using SBW system. The controlled variables are yaw rate and lateral acceleration, and the control input is front steering angle. On the other hand, to clarify the effect of the front steering angle controller, the reaction torque for adjusting steering feeling is set to be proportional to the steering wheel angle, with velocity-dependent gain, instead of a complicated vehicle motion feedback reaction torque [19,22].

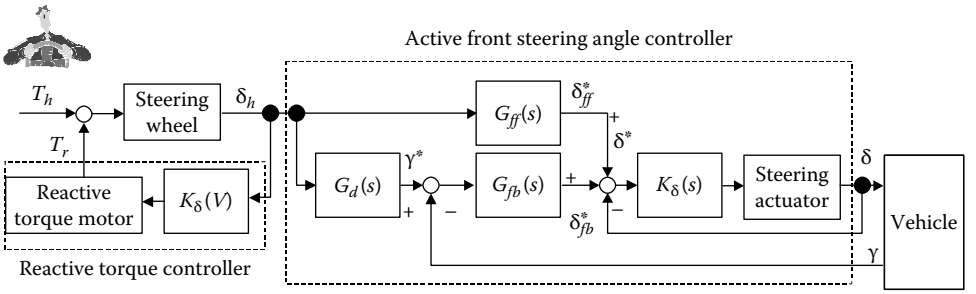


FIGURE 7.13 Block diagram of steer-by-wire control system.

The control system consists of a feedforward compensator, $G_{ff}(s)$ with respect to the steering angle, δ_h, and a feedback compensator, $G_{fb}(s)$ depending on the state deviation as shown in Figure 7.13 [22].

The desired yaw-rate model is determined to be the yaw-rate response of the zero-sideslip 4WS vehicle. To realize this characteristic, the relationship between steering wheel and front steering angle must be dependent on vehicle velocity. Moreover, to enhance stability against disturbance, steering control algorithm of feedback compensator is proposed by using theory of disturbance observer. The steering control laws of the feedforward compensator and feedback compensator can be expressed, respectively, as follows:

Feedforward compensator:

$$G_{ff}(s) = \frac{\delta^*_{ff}(s)}{\delta_h(s)} = \frac{k_{\gamma d}(V)}{k_\gamma(V)}\left[1 + \frac{\tau_\gamma(V) - \tau_{\gamma d}(V)}{\tau_{\gamma d}(V)s + 1}s\right] \tag{7.16}$$

Feedback compensator:

$$\delta^*_{fb} = \frac{1}{\tau_{fb}s + 1}\left[\frac{\tau_\gamma(V)s + 1}{k_\gamma(V)}\gamma - \delta_h\right] \tag{7.17}$$

where

$k_{\gamma d}$ and $\tau_{\gamma d}$ are the steady-state gain and the time constant of desired yaw-rate response, respectively

V indicates the chassis velocity

From earlier equations, the front steering angle can be expressed as follows:

$$\delta^* = \frac{k_{\gamma d}(V)}{k_\gamma(V)}\left[1 + \frac{\tau_\gamma(V) - \tau_{\gamma d}(V)}{\tau_{\gamma d}(V)s + 1}s\right]\delta_h(s) + \frac{1}{\tau_{fb}s + 1}\left[\frac{\tau_\gamma(V)s + 1}{k_\gamma(V)}\gamma - \delta_h\right]. \tag{7.18}$$

The other system was also proposed the alternative design of SBW control system by utilizing the combination of lateral acceleration and yaw rate, which is called D^* [19], for improving the vehicle stability, and examined the effectiveness of the control system by using experimental vehicle as shown in Figure 7.14.

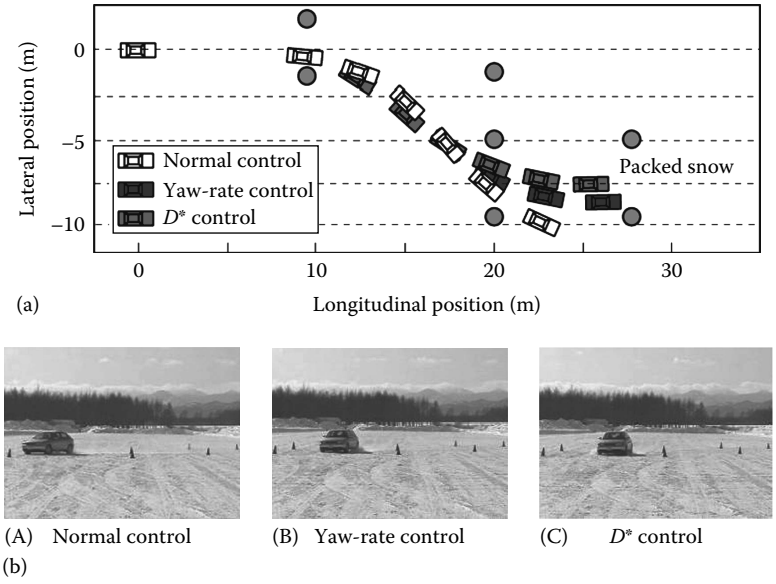


FIGURE 7.14 Results of full braking test on μ-split road. (From Shino, M. et al., *The Seventh International Symposium on Advanced Vehicle Control (AVEC '04)*, Arnhem, the Netherlands, pp. 797–802, 2004.)

7.5 SENSORS FOR AUTOMOBILE CONTROL

7.5.1 Sensing

The selection and the use of sensors are important matter for control and experiments. Sensors can translate status and attributes of physical world into values [25]. Appropriate sensors should be selected according to the observation objects, that is, required performance such as accuracy, response characteristics and also required frequency bandwidth. For example, the upper frequency of vehicle maneuver is ∼5 Hz for motion control. Then, the bandwidth of the sensors should be over 8 Hz.

In general, most sensors fall into analog and digital sensors [25]. Output signal from the sensors are gathered, processed, and handled using processors today [26]. Consequently, signal processing techniques are applied, that is, analog digital conversion and filtering [25]. Otherwise, in case of digital signal processing, analog data with sufficient sampling rate should be required to prevent an aliasing problem. An anti-aliasing filter is used in various types of digital signal processing methods. The flow of the sensing process is shown in Figure 7.15.

In the following section of this chapter, characteristics and the principle of sensing in several sensors for vehicle control and experiments are shown from the viewpoint of detecting own vehicle motion and vehicle surrounding condition.

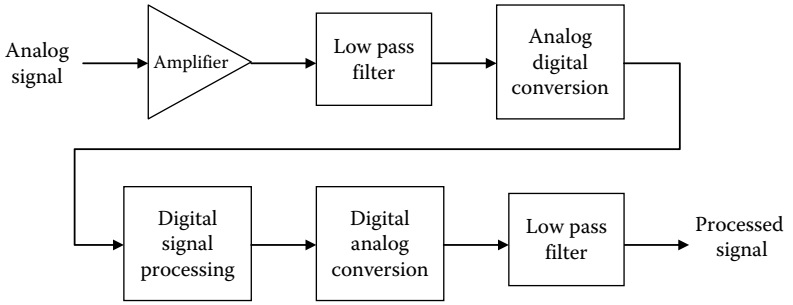


FIGURE 7.15 Sensing process.

7.5.2 Sensors for Detecting Own Vehicle Condition

7.5.2.1 Acceleration Sensor

For the vehicle control such as ABS and active suspension system, the detection of own vehicle condition is necessary. The acceleration sensors are classified to several types by the detecting theory as follows. Piezoelectric-type sensor consists of a PZT (lead zirconium titanate) or quartz as a piezoelectric element, a mass, and a housing. The PZT or quartz is sandwiched in the mass and housing. When it has displacement, the voltage can detect according to the piezoelectric effect. The features of this type sensor are compact, lightweight, and high sensitivity (Figure 7.16).

Strain gauge–type sensor's basic structure is same as the piezoelectric-type sensor. The piezoelectric element can be replaced to the strain gauge. It can measure the static acceleration [27]. If the beam is more rigid by spread the upper limit of measurement frequency, displacement will be reduced and sensitivity will be decreased.

Servo-type sensor has developed for inertial guidance of rockets. In the magnetic field, the moving coil moves according to the acceleration and measures the displacement. Current is sent to keep the displacement zero and that current is in proportional to the acceleration. The bandwidth is from DC to several hundred hertz. Resolution is very precise.

Oscillating-type sensor consists of a mass, which is supported by two strings. When acceleration a is given to the mass m, the tension of the left-side string will increase $\Delta T = ma$ and right-side one will decrease. Both strings are self-induced, vibrated, and mixing measured output frequency and natural frequency. Beat frequency is measured as the following formula:

$$f = \frac{1}{dL}\left(\sqrt{\frac{T+\Delta T}{\rho}} - \sqrt{\frac{T-\Delta T}{\rho}}\right) \tag{7.19}$$

where

dL is length of the string

ρ is the mass per unit length

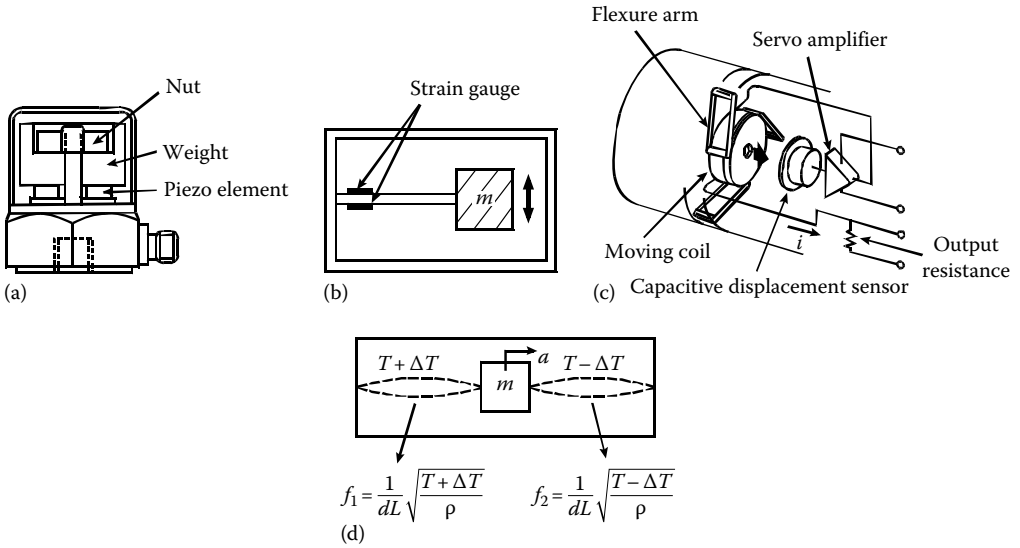


FIGURE 7.16 Acceleration sensors. (a) Piezoelectric-type sensor, (b) strain gauge–type sensor, (c) servo-type sensor, (d) oscillating-type sensor.

7.5.2.2 Yaw-Rate Sensor

For the VSC system, it is required to detect yaw motion of the vehicle, as well as the lateral and longitudinal acceleration and wheel rotating speed. The yaw-rate sensor has various methods, that is, gyro type, optical type, gas rate type, and so on [27]. One of the commonly utilized types is the vibrating structure type gyro, which is known as an inexpensive, small, and durable device. Figure 7.17 shows the basic principle of the vibrating structure type gyro. The arm inside the sensor is vibrated by a small sensor like a piezoelectric actuator. The vibrating object tends to keep vibrating in the same plane, when the sensor is rotated; therefore, the force on the object is produced according to the Coriolis effect. The sensor output voltage is in proportion to the force; therefore, it is possible to detect yaw rate by measuring the Coriolis force [27].

For the other type of gyro sensor, the ring laser gyro and optical fiber gyro are used. This type has higher performance in the stability and accuracy.

7.5.2.3 Force and Torque Sensor

For vehicle control and experiments, the measurements of force and torque are important. In the force sensor, there is a way of conversion to force from the elastic deformation, the amount of the electromagnetism, etc. [25]. Therefore, the force sensor is widely used as a basic element of other physical amount sensors like weight, pressure, and so on. The wheel force dynamometers can be used on test vehicles, servo-hydraulic and roller test stand. The piezoelectric-type sensor is used for a force sensor. It produces electric charges by the force applied. The piezoelectric elements are layered to detect the different directional forces and moments separately [28].

Steering torque is detected by the torsion angle of the torsion bar, which is located on the steering shaft differential transmission, and potentiometer and variable inductance methods are proposed [25].

7.5.3 Sensors for Detecting Vehicle Environment

7.5.3.1 Millimeter-Wave Radar

Laser radar has been in the market since 1995; however, it was not so robust in the rain condition. Radio wave radar has been developed to solve this problem, although it has problems about downsizing and reliability. The millimeter-wave radar measures a distance, a relative velocity, and direction between the sensor and the object.

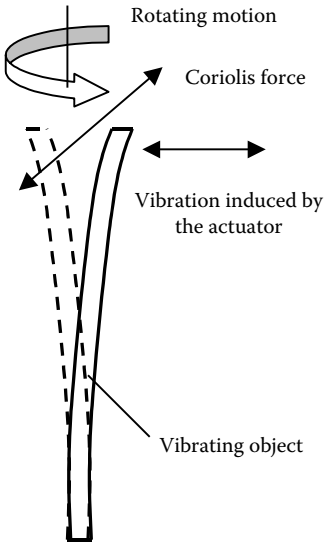


FIGURE 7.17 Principle of a vibrating structure type gyro.

The radar device consists of an antenna, a module, and a signal processing unit. The radar beam is generated and transmitted by the radio frequency (RF) module and the antenna. Frequency modulation (FM), pulse Doppler modulation, frequency-modulated continuous wave (FM-CW), spread spectrum radar wave system were used. The signal processing unit measures the delay of the transmitted and received wave. It also uses the FFT processing to find the beat frequency. To detect the direction of the object, the radar beam is scanned mechanically and electrically.

7.5.3.2 Laser Radar

Laser radar consists of a laser diode, a polygon mirror that is to scan the laser beam, and a photodiode which is to detect the reflect laser beam from the object and a CPU. It measures the flight time of the laser beam reflected from another vehicle's reflector. The CPU calculates the relative distance, velocity and estimates whether the vehicle is on the same lane or not. The rotation of the polygon mirror and pan-tilt angles is controlled by the DC motor.

7.5.3.3 Sonar

Sonar sensor uses ultrasonic wave to measure the distance. When transmitting the ultrasonic wave from the sensor, if there is some object, the sensor can receive the reflected wave. The distance is calculated by the flight time of the ultrasonic wave. The sensor consists of a ultrasonic microphone which has piezoelectric element and an amplifier. To transmit the ultrasonic wave, ultrasonic frequency AC voltage is sent to the microphone. The microphone vibrates according to the frequency of AC input. When the microphone receives the reflected wave, microphone generates the voltage according to the frequency of ultrasonic wave. The microphone uses as both transmitter and receiver.

7.5.3.4 Vision Sensor

Vision system consists of a camera and an image processor. Charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) camera is used [25]. To realize the real-time processing, high-performance image processor is used. High frame rate, 25 fps or more, is required.

It is used to detect the feature object, for example, lane marking, signboard and human shape for lane curvature prediction, collision avoidance, and so on [29]. However, it is difficult to use in poor light and bad weather condition. Then, the robust algorism development is the most important issue. It makes an attempt to use with the range finder and develop a sensor fusion system.

ACKNOWLEDGMENTS

Prof. Shigeyuki Nakadai (Chiba Institute of Technology), Prof. Taichi Shiiba (Meiji University), Prof. Kimihiko Nakano (The University of Tokyo), Prof. Motoki Shino (The University of Tokyo), Prof. Yohei Michitsuji (Tokyo University for Agriculture and Technology), Prof. Yoshiyuki Takahashi (Toyo University), and Dr. Shoichiro Takehara (The University of Tokyo) gave a lot of help for this chapter. The author would like to thank them for their great help.

REFERENCES

1. Y. Furukawa et al. A review of four-wheel steering studies from the viewpoint of vehicle dynamics and control, *Vehicle Syst. Dyn*. 18 (1989), 151–186.
2. T. Ehret et al. Control aspects of the Bosch-VDC, *The Third International Symposium on Advanced Vehicle Control (AVEC '96)*, Aachen, Germany (1996), pp. 573–608.
3. Y. Shibahata et al. The improvement of vehicle maneuverability by direct yaw-moment control, *The First International Symposium on Advanced Vehicle Control (AVEC '92)*, Yokohama, Japan (1992), pp. 452–457.
4. M. Abe et al. Direct yaw moment control with estimating side-slip angle by using on-board-tire-model, *The Fourth International Symposium on Advanced Vehicle Control (AVEC '98)*, Nagoya, Japan (1998), pp. 431–436.

5. M. Shino, P. Raksincharoensak, M. Kamata, and M. Nagai. Side slip control of small-scale electric vehicle by DYC, *The Dynamics of Vehicles on Roads and on Tracks*, Atsugi, Japan (2004), *Vehicle Syst. Dyn*. 41(Suppl.), 487–496.
6. M. Segawa, K. Nishizaki, and S. Nakano. A study of vehicle stability control by steer by wire system, *The Fifth International Symposium on Advanced Vehicle Control (AVEC 2000)*, Ann Arbor, MI (2000), pp. 233–239.
7. S. Kleine and J.L. Niekerk. Modeling and control of a steer-by-wire vehicle, *Vehicle Syst. Dyn*. 29 (1998), 114–142.
8. H. Winner, S. Witte, W. Uhler, and B. Lichtenberg. Adaptive cruise control systems aspects and development trends, SAE Paper No. 961010 (1995).
9. Y. Harada, H. Miyata, Y. Hayakawa, and S. Fujii, Cruise control system using adaptive control theory, SAE Paper No. 931917 (1993).
10. Y. Furukawa and M. Abe. Advanced chassis control systems for vehicle handling and active safety, *Vehicle Syst. Dyn*. 28 (1997), 59–86.
11. M. Nagai, Y. Hirano, and S. Yamanaka. Integrated control of active rear wheel steering and direct yaw moment control, *Vehicle Syst. Dyn*. 27 (1997), 357–370.
12. H. Takano and M. Kato. *Technologies of Actuators*. Riko-Gakusha, Tokyo, Japan, 1999 (in Japanese).
13. K. Nakano and Y. Suda. Combined type self-powered active vibration control of truck cabins, *Vehicle Syst. Dyn*. 41(6) (2004), 449–474.
14. M. Amano and T. Nogi. Application of electrical and electronic technologies to automotives, *J. Soc. Automot. Eng. Jpn*. 59(5) (2005), 4–9 (in Japanese).
15. A. Shimada, H. Ogawa, and M. Nakajima. Development of an ultra-thin DC brushless motor for a hybrid car, *JSAE Rev*. 22 (2001), 287–292.
16. Y. Suda, K. Suematsu, K. Nakano, and T. Shiiba. Study on electromagnetic suspension for automobiles—Simulations and experiments of performance, *The Fifth International Symposium on Advanced Vehicle Control (AVEC 2000)*, Ann Arbor, MI (2000), pp. 699–704.
17. D. Karnopp et al. Vibration control using semi-active force generators, *Trans. ASME, J. Eng. Ind*. 96(2) (1974), 619–626.
18. M. Segawa et al. *Development of Steering Hardware-in-the-Loop Simulator*, FISITA World Automotive Congress, Yokohama, Japan (2006), F2006V196.
19. M. Segawa et al. Vehicle stability control strategy for steer-by-wire system, *JSAE Rev*. 22 (2001), 383–388.
20. Y. Terashima et al. Development of a micro vehicle for severely disabled people, *The Fifth International Conference on Motion and Vibration Control (MOVIC)*, Saitama, Japan, Vol. 2 (2002), pp. 855–860.
21. T. Hosokawa et al. Enhancing maneuverability with driving manipulating devices based on older driver physical characteristics, *JSAE Rev*. 24 (2003), 359–361.
22. M. Shino et al. Vehicle handling and stability control of micro-scale electric vehicle utilizing steer-by-wire system, *The Seventh International Symposium on Advanced Vehicle Control (AVEC '04)*, Arnhem, the Netherlands (2004), pp. 797–802.
23. C.D. Gadda et al. Eliminating the need for sensor redundancy in diagnostic system for steer-by-wire vehicles, *The Eighth International Symposium on Advanced Vehicle Control (AVEC '06)*, Taipei, Taiwan (2006), pp. 59–64.
24. Y.H.J. Hsu et al. A feel for the road: A method to estimate tire parameters using steering torque, *The Eighth International Symposium on Advanced Vehicle Control (AVEC '06)*, Taipei, Taiwan (2006), pp. 835–840.
25. K. Tada et al. *Sensor Technology*, Maruzen Advanced Technology, Maruzen, Tokyo, Japan (1991), pp. 2, 15, 35, 51–58, 228.
26. M. Hatori et al. *Digital Signal Processing*, Maruzen Advanced Technology, Maruzen, Tokyo, Japan (1994), p. 17.
27. H. Miura et al. *Handbook Mechatronics*, Ohmsha, Tokyo, Japan (2005), pp. 153–157.
28. Kistlar: Report SD920-234e-11.02.
29. P.A. Liatsis et al. Novel lane support framework for vision-based vehicle guidance, *Industrial Technology, 2003 IEEE International Conference*, Maribor, Slovenia, Vol. 2 (2003), pp. 936–941.

8 Optimization of Ground Vehicle Systems

Massimiliano Gobbi and Panos Y. Papalambros

CONTENTS

8.1 Introduction 241
8.2 Basics 242
8.2.1 Formulation of the Optimization Problem 242
8.2.2 Solution Methods 244
8.2.2.1 Nonlinear Programming 245
8.2.2.2 Search Based on Uniformly Distributed Sequences 245
8.2.2.3 Evolutionary Algorithms 245
8.2.2.4 Global Approximation 246
8.2.2.5 Optimal Design of Multilevel Systems 246
8.2.2.6 Comparison of the Different Solution Methods 247
8.2.3 Reliability-Based Design Optimization 248
8.2.4 Analysis of the Pareto-Optimal Set 248
8.2.5 Available Commercial Software for Engineering Design Optimization 249
8.3 Optimal Design of Road Vehicles and Subsystems 250
8.3.1 Vehicle System Dynamic Behavior 250
8.3.2 Power-Train Design 251
8.3.3 Internal Combustion Engine Design 251
8.3.4 Safety and Ride Comfort 252
8.3.5 Vehicle System Design and Lightweight Structures 252
8.3.6 Integration of Vehicle Electronic Controls 253
8.4 Optimization of the Dynamic Behavior of Vehicles 254
8.5 Conclusion 256
References 256

8.1 INTRODUCTION

The design process is and will always be based on the designer intuition. However, optimization techniques can permit significant improvement in the process itself. The designer activity undergoes iterations between an analysis stage, in which the relationships between parameters, performance indices, and design constraints are assessed, and a synthesis stage, in which the possible and desired performances are obtained by a proper parameter setting. The inherently iterative nature of the design process makes it particularly difficult to keep it under control.

The application of optimization in automotive engineering is increasing as more powerful computers become available. Today, calculations can be performed in a fraction of the time it took a few years ago. Therefore, the applications for numerical optimization have increased.

It is assumed in the following that an accurate (validated) mathematical model of the system under optimization is available. Successful optimization requires availability of appropriate

analysis models and knowledge of the capabilities and limitations of the mathematical optimization techniques [6,79,88,95].

In real-world applications, problems with a single, well-defined objective function tend to be the exception rather than the rule. In engineering, most optimization problems are multi-objective in their nature. In a problem with multiple objectives, it is generally impossible to obtain a ranking of all of the alternative solutions, without using further rules or assumptions. In this situation, the decision of the better solution may become subjective, or must rely on additional information, such as the importance of each objective. The most common and rigorous approach to overcome this limitation is to define the efficient solutions set [9,43,81]. If possible, the designer has to identify hierarchical levels in the complex system under study so the system can be divided into subsystems that can be further subdivided. Analytical target cascading (ATC) is a methodology for optimal design of hierarchical multilevel systems where outputs of lower-level elements are the inputs to higher-level elements [67,70,97]. In any case, optimization of complex systems may require a prohibitive simulation effort, even with the most efficient techniques. In this case, the relationship between design variables and objective functions is expressed by means of a global approximation approach. The physical model is substituted by another purely mathematical model, which is able to give computational results very quickly [33,49,82]. The inclusion of uncertainty in the problem formulations is a relatively recent development in optimization [8,10,34,36,113]. The approach allows the designer to minimize the objective function while accounting for the influence of stochastic variations of some parameters and/or variables. Typically, this involves trading (deterministic) optimality for robustness, whether in the value of the objective (robust design) or in the satisfaction of the constraints (reliable design). After the optimal solutions have been computed [12,49,81,105], the designer can make a choice and select from this set the preferred solution featuring the desired compromise among objective functions [22].

A brief outline of optimal design methods generally used for vehicle design problems is given later. Some examples of design optimizations will be given in the field of vehicle system dynamics, active safety and ride comfort, vehicle system design and lightweight structures, and advanced automotive electronics [9,27,31,43,67,83,100].

8.2 BASICS

In this section, some basic concepts of optimization theory are reviewed. The interested reader may refer to [6,79,88,95,96] for an in-depth documentation on these topics. It is assumed that an accurate (validated) mathematical model of the system under optimization is available.

8.2.1 Formulation of the Optimization Problem

The majority of engineering problems involve a constrained optimization formulation, that is, the task of minimizing (or maximizing) a vector of k objective functions (performance indices) subject to different types of constraints. Objective functions and constraints must be quantifiably expressed as functions (f, g) of the design variables x.

Find

$$\begin{aligned} &\min_{x\in R^n} f(x) \\ &h_j(x) = 0, \quad j = 1, \ldots, m_1 \\ &g_j(x) \le 0, \quad j = 1, \ldots, m_2 \\ &x \in X \end{aligned} \tag{8.1}$$

where

f is the objective function vector
x is the vector of the design variables
X is the definition domain of x
h_j and g_j are equality and inequality constraints, respectively

The set X can represent certain ranges of real values or certain types, such as integer or "standard" values, which are very often used in design specifications [88,95]. Should a maximization problem occur (i.e., max $f(x)$), the vector function $f(x)$ could be rewritten as $-f(x)$ or $1/f(x)$ and formulation 8.1 could be still used [6].

The most common and rigorous approach to solve a multi-objective optimization problem is to define *efficient solutions* (and then select) as those (and only those) solutions that improve at least one objective function and worsen at least one of the others. Under this assumption, the set of efficient solutions constitutes the pareto-optimal set [72,75,79]. The disadvantage of this approach is that generating pareto-optimal sets is generally computationally expensive.

The pareto-optimal set is the set of the non-dominated solutions, that is, the thick line in Figure 8.1 for a problem with two design variables and two objective functions. In this case, the set still represents a trade-off between the two objectives f_1 and f_2. Each point in the rectangle on the plane $x_1 - x_2$ (design variables domain) corresponds to a point in the manifold on the plane $f_1 - f_2$ (objective functions domain). The rectangular manifold on the plane $x_1 - x_2$ is transformed into the manifold on the plane $f_1 - f_2$ by the function $f(x_1, x_2)$. As all the possible combinations of design variables x_1 and x_2 have been used to generate $f_1(x_1, x_2)$, $f_2(x_1, x_2)$ (if the solution is compatible with the design constraints). Referring to Figure 8.1, point A is said to be dominated by B and C, that is, B and C dominate A. A is also dominated by all the design solutions between B and C on the bold line, that is, the solutions on the bold line between B and C are better at least in one objective function than A. Given a (wrong) solution A, there exist solutions that are better than A at least in one objective function. All the solutions corresponding to points in Figure 8.1, which do not lie on the pareto-optimal set, shown as bold line in Figure 8.1, have to be discarded by the designer. The pareto-optimal set for a biobjective problem is limited by the minima for the two scalar objective functions f_1 and f_2.

The pareto-optimal solutions are defined mathematically as follows. Given the minimization problem (8.1) with k objective functions and n design variables, a solution x_i is pareto-optimal, if there is no solution x_j such that

$$\begin{cases} f_m(x_j) \le f_m(x_i) \quad m = 1, 2, 3, \ldots, k \\ \exists l: f_l(x_j) < f_l(x_i) \end{cases} \tag{8.2}$$

This definition can be used to find the pareto-optimal solutions directly (pareto filtering, see Table 8.1) [81].

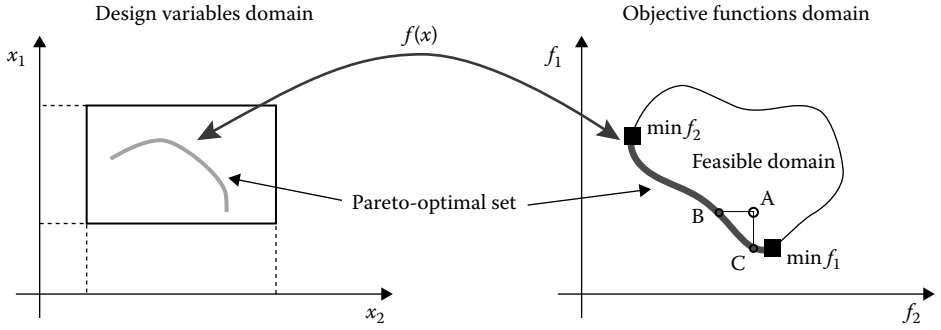


FIGURE 8.1 Multi-objective optimization problem with two design variables x_1, x_2 and two objective functions f_1, f_2. Feasible set and pareto-optimal set in the design variables and objective functions domains.

TABLE 8.1
Overview of Multi-Objective Optimization Methods

Scalarization Methods	Pareto Methods
Weighted sum (linear and nonlinear)	Pareto filtering
Constraints method	Adaptive weighted sum method
Multi-attribute utility analysis	Normal boundary intersection
Goal programming	Multi-objective evolutionary algorithms

Several methods exist for converting the multi-objective formulation into one with a scalar objective function that can be solved with the usual single-objective optimization methods, typically using some sort of weighting function. The shortcoming of such scalarization is that the designer must include subjective information and preferences a priori. Additionally, some optimal solutions can be unattainable for non-convex problems (see [88,95]).

Table 8.1 provides a list of multi-objective optimization methods. An overview of these methods can be found in [19,79]. Some scalarization methods can also be used to generate the pareto-optimal set [20].

The multi-objective optimization problem (Equation 8.1) can be reformulated ("weighted sum method") as a single-objective problem where the weighted sum of the objective functions has to be minimized. By changing the weight settings, different pareto-optimal points can be generated. This approach is usable only if the pareto-optimal set is convex.

By applying the "constraints method," the multi-objective optimization problem (Equation 8.1) with k objective functions can be converted into a single-objective problem by minimizing one objective function and adding inequality constraints (upper bounds) to the remaining $(k - 1)$ objective functions.

8.2.2 Solution Methods

The problem in Equation 8.1 is formally a mathematical programming problem [6,88,95,96]. The functions f and g in Equation 8.1 can be expressed with algebraic equations or computer simulations. If the functions f and g are all linear, the problem is a linear programming one. Otherwise, the problem is a nonlinear programming one. Discrete programming refers to problems where all the design variables have only discrete values. A large class of design problem comprises mixed-discrete/continuous design variables values [96]. This type of problems is usually a challenge for the standard methods of nonlinear programming. In Table 8.2, some of the methods for solving optimization problems are presented and compared [77,79]. This topic is further discussed in the following subsections.

TABLE 8.2
Methods for Solving Optimization Problems

Method	Type	Efficiency	Accuracy		Discrete Design Variables
Exhaustive	Nonderivative	– –	++ (Discrete variables)	– (Continuous variables)	Yes
Uniformly distributed sequences	Nonderivative	–		–	Yes
Evolutionary strategies	Nonderivative	+		+	Yes
Sequential quadratic programming	Derivative	+ +		+	No

Note: – –, very bad; –, poor; +, good; + +, very good.

8.2.2.1 Nonlinear Programming

Nonlinear programming algorithms have been extensively studied in the literature.

These algorithms are able to solve single-objective optimizations by means of an iterative procedure starting from an initial guess.

On the basis of the order of the method, three main classes can be considered:

- *Zero-order methods*: The new solution is computed without using any information about the first derivatives. Among the others, there are the random walk, the pattern search, and the simplex method [96].
- *First-order methods*: The iteration step is computed by using the gradient that identifies the steepest descent direction. In this way, a reduction in the objective function is guaranteed at each step. The steepest descent method and the conjugate gradient method [95] belong to this class.
- *Second-order methods*: The curvature (Hessian matrix) is used along with the gradient. These methods are also known as Newton's methods [6].

The zero-order methods may be useful when the functions are not differentiable. In the other cases, gradient-based methods are widely used as the Hessian computation might be too demanding. By numerical differentiation (finite differences), the computational cost needed to estimate the gradient and the Hessian matrix is proportional to n and n^2, respectively, for a function of n variables. For this reason, the quasi-Newton methods have been developed, where the Hessian matrix is estimated by means of the function values and the gradient values.

8.2.2.2 Search Based on Uniformly Distributed Sequences

Within the large set of methods to find optimal solutions, the *exhaustive search method* [81] is the simplest, but it requires a huge amount of computational effort. Orthogonal arrays and low discrepancy sequences are used to reduce as much as possible the number of combinations of design variables used to explore the design variable domain [109]. In multi-dimensional spaces (more than 3 and up to 30 or 40), a low discrepancy sequence can be used to define a set of design variables combinations to be exploited to sample the objective functions domain. Low discrepancy sequences differ from pseudo-random sequences in that the points are more evenly distributed in the feasible space (see [81,109] for a detailed description of this topic).

8.2.2.3 Evolutionary Algorithms

Evolutionary strategies are a class of non-gradient methods, which has grown in popularity since Holland first published his book on the subject [53]. Two types of evolutionary algorithms are considered here, namely, genetic algorithms (GAs) and simulated annealing (SA).

GAs are suitable for finding the minimum of a function (or of a set of functions) by performing a semi-stochastic search [12,17,46,53,99,105]. The design variable vector referring to a particular design solution is typically encoded into binary or real strings (chromosomes). GAs are based on an elitist reproduction strategy, where the strongest members of the population (design solutions) are selected for reproduction and are given the opportunity to strengthen the chromosomal (i.e., genes, namely, design variables) makeup of the next generation. Unlike many other search techniques, GAs consider multiple design solutions (a population) at each iteration.

Some comments on GAs are as follows:

- GAs work on function evaluations alone and do not require function derivatives.
- GAs proceed from several points in the design variable domain (population); consequently, the method has a better probability of locating the global minimum.
- GAs allow design variable spaces consisting of a mix of continuous and discrete variables.

- GAs use probabilistic transition rules, not deterministic rules.
- GAs can be easily implemented on parallel computers.
- Computational costs are high, and good performance often requires tuning of the algorithm for the specific problem at hand.

A ranking method can be easily included to grade the population in terms of pareto-optimality and construct a procedure able to identify the entire pareto-optimal set and not, as it occurs for standard algorithms, one single pareto-optimal solution for each run [12,18,46,105].

SA is a heuristic search method for obtaining good solutions to difficult optimization problem. Kirkpatrick and colleagues [65,96] showed how a model of simulating the annealing process in solids as proposed by Metropolis could be used for optimization problems.

Metropolis et al. [80] introduced a simple algorithm to simulate annealing process by considering a collection of atoms at a given temperature (T). To apply Metropolis algorithm for other than thermodynamic process, one must provide the following:

1. A description of possible system configurations
2. A generation of the random changes in the configurations
3. An objective function, whose minimization is the goal
4. A control parameter T and an annealing schedule

There has been some work done in multi-objective SA where the non-dominated solutions found during search are stored during the annealing process [107].

8.2.2.4 Global Approximation

Complex system design requires extensive use of simulation-based design and analysis tools, such as multibody dynamics simulations (MBD), finite element analysis (FEA), computational fluid dynamics (CFD), and other computationally intensive models. The need to quantify the engineering performance of complex products, such as automobiles, leads to computationally expensive models due to fidelity of modeling and to the obvious need for accurate results. Moreover, computer simulation codes can be discipline-specific, distributed, on different computer platforms. Sometimes, computationally expensive computer simulations and/or analyses are replaced by surrogate models, which are fast, simple approximations. To approximate functions, a vast series of techniques exist that go from a simple polynomial interpolation to very more accurate methods of representing function behavior, such as radial basis neural networks or Kriging models [97].

Optimization may require a prohibitive simulation effort, even with the most efficient techniques. In this case, the relationship between design variables and objective functions can be *globally approximated* by means of a simpler, approximate mathematical model (e.g., an artificial neural network or a piecewise quadratic function or other approximation). This approximate or surrogate model or pure numerical model is defined on the basis of a limited number of simulations (i.e., computational experiments) performed by means of the originally validated physical model of the system under consideration.

Typically, the simulation time for the pure numerical model is a small fraction (1/10 to 1/10,000) of the simulation time needed by the original model [82,85].

The accuracy in the approximation of the outputs depends on the approximation model employed and the computational effort expended to build the model. The first attempts using neural networks for global approximation were reported in [49]. A comparison of different approximation methods is given in [33] and some results are shown in Table 8.3.

8.2.2.5 Optimal Design of Multilevel Systems

Every system is analyzed at the level of complexity that corresponds to the interest of the designer. For this reason, we can identify hierarchical levels in the system definition; the system can be divided into subsystems that can be further subdivided.

TABLE 8.3
Approximation Methods

		Evaluation Accuracy		
	Approximation Domain	**Simple Model**	**Complex Model**	**Tuning Effort**
Linear interpolation	Local	+	−	+ +
Quadratic interpolation	Local	+	−	+ +
Neural networks (multilayer Perceptron)	Global	+ +	+ +	−
Statistical approximation (Kriging)	Global	+ +	+	−

Note: −, bad; +, good; + +, very good.

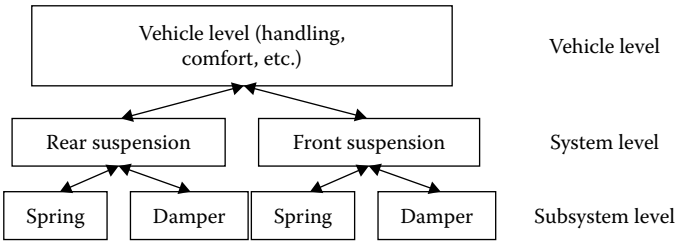


FIGURE 8.2 Example of hierarchical multilevel system decomposition.

ATC is a methodology for optimal design of hierarchical multilevel systems where outputs of lower-level elements are the inputs to higher-level elements (see Figure 8.2).

The objective is to identify the element interactions early in the design process, and to determine specifications that yield consistent system design with minimized deviation from design targets. The process formulates and solves a minimum deviation optimization problem for each element of the multilevel system. By solving the element optimization problems sequentially in an iterative manner, ATC aims at minimizing discrepancies between the optimal design variable values desired at elements higher in the hierarchy and response values that elements at lower levels can actually deliver [64,68,86,87,95]. In addition, if design variables are shared among some elements at the same level, their required common final optimal value is coordinated by their parent element at the level above.

8.2.2.6 Comparison of the Different Solution Methods

There is no unique answer to which optimization method is the best. Many studies have shown that different methods show the best performance, depending on the problem and on how much effort has been spent to tune the method itself (the procedure itself needs to be optimized).

Search based on uniformly distributed sequences can be computationally very expensive, but no parameters relative to the algorithm itself need to be tuned. GAs are robust and they can handle multimodal functions, but several parameters (population size, mutation, crossover probability, etc.) need to be tuned to get adequate convergence properties. SA is a robust method slightly less computational expensive than GA.

Standard nonlinear programming methods show very fast convergence rate and they are very easy to be tuned, but they are not as robust (i.e., get stuck in local optima) as the other methods. Moreover, they can just find one optimal solution for each run. Moreover, the solution of engineering

problems often requires a mixture of numerical simulations, analytical calculations, and selection of catalog components. Therefore, non-gradient optimization methods (i.e., search based on uniformly distributed sequences, GA, SA) are better suited for these types of problems, being impossible the calculation of the derivatives of the objectives function.

8.2.3 Reliability-Based Design Optimization

Because deterministic optimum designs obtained without taking uncertainty into account could lead to unreliable designs, a reliability-based approach to design optimization is preferable. Recently, many authors have introduced procedures that combine design optimization techniques with reliability-based design methods [3,8,41,45,69,74,103,104,113,114]. The approach allows the designer to minimize the objective function while accounting for the influence of stochastic variations of some parameters and/or variables. Typically, this involves trading (deterministic) optimality for robustness, whether in the value of the objective (robust design) or in the satisfaction of the constraints (reliable design).

In the presence of random variables and parameters, responses, such as objective and constraint value calculations, are themselves random variables, whose expected values and/or variances must be computed when solving probabilistic optimization problems. Evaluation of probability distributions of responses can be achieved using Monte Carlo simulation, but this is a very expensive process. Thus, various approximation schemes must be used in practical situations.

A mean-value first-order second-moment approach (first-order Taylor expansion about the current design) can be adopted to estimate the mean and standard deviation of the objective functions [45]. When a trained artificial neural network is used to compute model responses, an accurate estimate of the variance of the objective functions can be obtained with no additional computational cost [34,45].

Uncertainty can also come from considering market behavior when launching new products. For example, developing a new vehicle design in a dual-use context (commercial enterprise and government military uses) is affected by uncertainty in market penetration and cannibalization [14].

8.2.4 Analysis of the Pareto-Optimal Set

After the optimal solutions have been computed (pareto-optimal set), the designer can make a choice and select from this set the preferred solution featuring the desired compromise among objective functions.

This process can be preceded and even accelerated by special analyses allowing the designer to have an insight into the physical phenomena he is trying to analyze. In other words, there are some analyses by which the designer may understand the reason why the considered optimal solution requires such a design variable combination. In particular, these analyses may show which is

- Relationship between two objective functions (within the pareto-optimal set)
- Relationship between design variables (within the pareto-optimal set)
- Relationship between design variables and objective functions (within the pareto-optimal set)

These different relationships are shown in Figure 8.3.

The Spearman rank-order correlation coefficient [43,79,92] can be used to assess the earlier introduced relationships. This rank-order correlation coefficient has proved to be robust in presence of nonlinearity or outlayers.

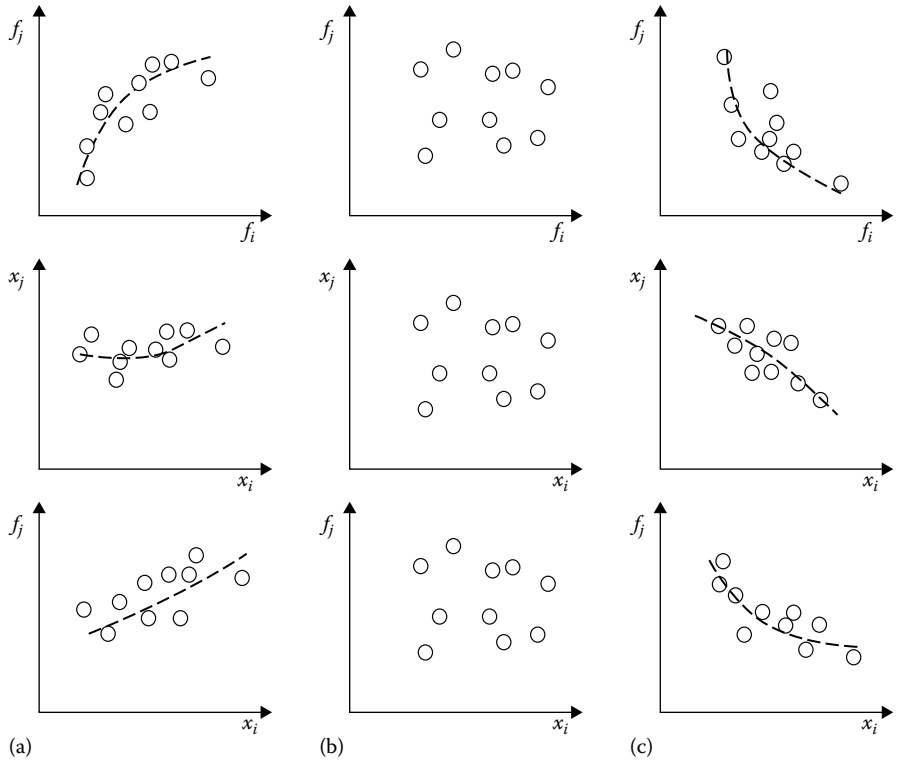


FIGURE 8.3 Relationship between pareto-optimal objective functions ($f_i - f_j$), between pareto-optimal design variables ($x_i - x_j$), and between objective functions and design variables ($f_i - x_i$). The values can be either directly correlated (column a), uncorrelated (column b), or indirectly correlated (column c).

8.2.5 Available Commercial Software for Engineering Design Optimization

A list of commercial software suitable for solving design optimization problems related to vehicles is given as follows.

The list is by no means exhaustive. The aim is instead to give to the reader an idea of the wide set of software solutions available on the market nowadays:

- ADAMS/Insight [59] is the optimization tool designed for use within the ADAMS multibody dynamics simulation environment. The software can use design of experiment techniques in order to define a global approximation of the objective functions (polynomial fit). The optimizer includes standard nonlinear programming algorithms and allows a multiobjective formulation of the design problem. Some analysis tools can be employed to identify the design variables that most affect the system design.
- FRONTIER [55] has a strong emphasis on multi-objective formulations. A design of experiments module can be used in the preliminary design phase (factorial, Sobol, etc.). Linear and nonlinear response surfaces for data modeling and approximation can be used (polynomial fitting, neural networks, etc.). A reliability-based design optimization tool is included.
- GENESIS [61] is a structural analysis and optimization software. It integrates various FEA and optimization capabilities (sizing, shape, topography, topometry, topology, etc.). It can also be used as a more general-purpose optimizer.
- iSIGHT [62] is a general-purpose optimization tool with a large number of alternative algorithms for optimization and building surrogate model functions. It includes ability to solve mixed-discrete problems and to generate pareto sets. It can also be used to perform probabilistic optimization.

- MATLAB® Optimization Toolbox [58] is a general-purpose collection of functions for optimization. The toolbox includes nonlinear programming, and multi-objective minimization algorithms (goal attainment, etc.). The tool can be integrated with different Toolboxes useful in building surrogate model functions. GA and Direct Search Toolbox extend the optimization capabilities with tools for using GA, SA, and direct search.
- NEXUS [57] is a general-purpose multidisciplinary and multi-objective optimization tool. The software can solve nonlinear, constrained, and unconstrained as well as mixed-discrete problems. A modular plug-in architecture extends its basic capabilities. Specific modules implement Design of Experiments and Response Surfaces methodologies. Standard nonlinear programming, direct search and evolutionary algorithms for single- and multi-objective optimizations are available.
- OPTIMUS [60] is a general-purpose optimization tool. The software can explore the design space using design of experiments; it includes several standard nonlinear programming algorithms and state-of-the-art evolutionary algorithms. Special algorithms are available for solving problems including a mixture of continuous and discrete variables. Multi-objective solution methods are implemented. A reliability-based design optimization tool is included.
- OPTISTRUCT [56] is a finite element–based software for both structural analysis and design optimization. The integrated state-of-the-art gradient-based optimization methods allow component size and shape optimization (sizing, shape, topography, topology, etc.). OptiStruct includes morphing technology to prepare finite element meshes for optimization. Size optimization defines ideal component parameters such as material type, cross-section dimensions, and thicknesses. Design of experiments capabilities and stochastic studies are included in the HyperStudy tool, which can be directly interfaced to OptiStruct.
- VisualDOC [61] is a general-purpose optimization tool. The software can perform linear, nonlinear, constrained, and unconstrained as well as integer, discrete, and mixed-discrete optimization. Gradient-based, non-gradient–based, and response surface approximate optimization algorithms are available. In addition, a design of experiments module and probabilistic analysis and design capabilities are included.

8.3 OPTIMAL DESIGN OF ROAD VEHICLES AND SUBSYSTEMS

The following state-of-the-art review is adapted from [32] and is reported here in an updated and extended form.

8.3.1 Vehicle System Dynamic Behavior

The first attempts to design optimally the suspension system of a road vehicle appeared in [16,52,54,63,76,89,102,111]. Some basic studies on the optimization of the ride comfort and active safety of road vehicles are presented in [37,39,40,90]. Both passively and actively suspended vehicles have been considered. Simple analytical formulae for the estimation of vehicle dynamic behavior are derived, and these formulae are used in conjunction with multi-objective programming (MOP) to find the optimal suspension design variables symbolically, ensuring the best compromise among comfort, road holding, and working space. In [36,37], the design variables, that is, suspension stiffness and damping, are considered as stochastic variables and the vehicle body mass and tire radial stiffness as stochastic parameters, in order to derive the optimal trade-off solutions in a optimal and robust framework.

Referring to the tuning of gains of an actively suspended vehicle, analytical formulae seem to have been derived only for simple system models (results reported in [2,40,54,110], establishing the

well-known "skyhook" control strategy). A basic study on the optimization of the handling behavior and active safety of road vehicles is presented in [83]. The work demonstrates how the front and rear tire cornering stiffness of a vehicle can be tuned in order to obtain a preferred compromise among the many conflicting performance indices that describe the handling behavior and active safety. Similar studies in optimizing vehicle handling behavior over a range of maneuvers but involving a much greater number of design variables are reported in [35,43,78,84,98,100,101,108]. In order to design a car effectively, one is often required to tune the entire set of design variables related to tires, aerodynamics, and chassis characteristics (stiffness, damping, and kinematics of the suspension system). Such tuning presents good opportunities for effective use of optimization methods [45].

8.3.2 Power-Train Design

Hybrid-electric propulsion systems have been optimized with respect to fuel economy, ride, and mobility targets in [67] by applying a target cascading approach.

A method for combined design optimization and power management of a hybrid propulsion system integrated within a complete vehicle system was proposed in [28]. A sequential quadratic programming (SQP) algorithm with a design space exploration–based multi-start technique was used to enhance the chances of achieving the global optimum.

The design of the synchronizer and selector mechanism of a vehicle gearbox is discussed in [41,42]. The main aim there is to improve shifting ability during a reference shift action. Fifty-eight variables of the system model were thus tuned. The approach proposed in [41] addresses both the optimization of nine system performance indices and the reduction of the sensitivity (variance) of the performance indices to stochastic perturbations. The variances are computed by means of a procedure based on the global approximation of the objective functions.

The use of ATC in vehicle transmission design was reported in [7]. ATC was applied to the simulation-based design of a CVT implemented in a mid-sized truck. The overall system was decomposed into three levels. The study highlighted two main issues. First, the development of the coordination strategy and appropriate models appeared to be an intricate task that benefits from extensive knowledge of the overall problem. Second, the values of the arbitrary weights defined at all levels had a significant impact on the results and of ATC as a practical and suitable approach.

8.3.3 Internal Combustion Engine Design

A probabilistic formulation of the ATC process has been used in [10,70] to solve an internal combustion engine design problem. An engine is considered at the top level system, which is then decomposed into subsystems representing the piston-ring/cylinder-liner subassembly of the cylinders. The system model predicts engine performance in terms of brake-specific fuel consumption. The ring/liner subassembly simulation takes as inputs the surface roughness of the ring and the liner (assumed to be normally distributed) and the Young's modulus and hardness and computes power loss due to friction. The engine simulation takes then as input the power loss and computes brake-specific fuel consumption of the engine.

An analysis of an optimized approach for designing engine mounts is given in [1]. GAs and SQP have been used for selecting the combination of parameters that can yield the deepest notch and shortest resonance peak in the dynamic stiffness of the fluid mounts.

The link between manufacturing process and product performance is studied in order to construct quantifiable criteria for the introduction of new engine technologies and processes in [73]. Cost associated with a new process must be balanced against increases in engine performance. A predictive engine simulation model is used to quantify performance gains due to the new surface finish obtained by means of abrasive flow machining for finishing the inner surfaces of intake manifolds for two engines. Subsequently, economic cost-benefit analysis is used to evaluate manufacturing decisions based on their impact on the firm's profitability.

8.3.4 Safety and Ride Comfort

In [5], GAs are used for determining the active control and passive mechanical parameters of a vehicle suspension system in order to minimize the vertical acceleration subject to constraints on road holding and suspension working space. In [31], a MOP approach using GAs is employed for the design of a multi-link suspension, based on performance indices related to vehicle handling and stability.

In [84], a procedure is presented for the integrated design (tuning) of tires and suspensions of a racing car. Proper objective functions are defined after a subjective–objective correlation analysis. Different driving situations (steady state, J-turn, lane change, power on–off while steering, braking on a bend, passing over a kerb while steering) are optimized with respect to 18 design variables related to the suspension system and the tire characteristics. A global approximation model was used together with a search method based on low discrepancy sequences. The same approach has been followed in [24,35,98,101] for the problem of the chassis design of a production car. In [43], the computation of the pareto-optimal set is performed, for a similar chassis design problem by using GAs. In [45], a method for the reliability-based multi-objective design optimization of tires and suspensions of a racing car is introduced and applied. The optimization has been performed on the basis of a global approximation procedure.

In [47], through the use of an optimization algorithm with finite-difference sensitivities, the suspension characteristics have been optimized. The ride optimization is achieved by finding the optimum of a ride index that results from a metric that accounts for the acceleration in several key points in the vehicle properly weighted. The extension by including the optimization of the handling characteristic is given in [48].

The design of heavy vehicles for good dynamic performance in a variety of scenarios or maneuvers (from standard test maneuvers to extreme emergency maneuvers) is discussed in [9]. The problem is formulated as a multi-criteria, multi-scenario design problem with the goal of finding an optimal vehicle design, which can improve the vehicle dynamic performance in all the considered scenarios simultaneously (lane change, pulse steer, and ramp steer, induced rollover). A modified Monte Carlo optimization technique is used to find the optimal designs.

A robust design approach for improving vehicle crashworthiness was applied using the Monte Carlo simulation method on a global response surface in [50].

In [103], a methodology for reliability-based multi-objective design optimization of automotive body components under impact scenarios is introduced. The approach is applied to minimize the structural mass while maximizing energy absorption of an automotive rail section under structural and occupant safety constraints.

In [66], the focus is on design optimization for safety and NVH. The set of vehicle system design variables is tuned in order to minimize the vehicle weight while satisfying constraints on the static torsion and bending displacements and limiting the injury criteria during frontal crash and side impact.

ATC has been used to optimize ride quality and handling performances by considering suspension, tire, and spring analysis models in [64,68,69]. Potential incompatibilities among targets and constraints throughout the entire system can be uncovered and the trade-offs involved in achieving system targets under different design scenarios can be quantified.

The problem of choosing a single solution within a set of pareto-optimal alternatives is addressed in [72]. Two methods, the k-optimality approach and the more general $k\varepsilon$-optimality method, were considered. These two methods theoretically justify and mathematically define the designer tendency to choose solutions that are "in the middle" of the pareto-optimal set. The methods were successfully applied to the optimization of the tire/suspension system of a car.

8.3.5 Vehicle System Design and Lightweight Structures

The optimal layout of structural components is discussed in many papers [38,51,71,75]. The problem of the definition of the layout of the entire vehicle is addressed in recent papers [51]. Real-coded GAs have been used to solve the multi-objective optimization problem employing an elitist search.

An efficient way to find good lightweight designs for vehicle structures is to use different tools for structural optimization [4,21,27,30,93,106]. Structural optimization includes several approaches, for example, material, size, shape, and topology optimization, where size and shape optimization are used for fixed topologies and for improvements of existing structures. Topology optimization, however, is a more general optimization method that can be used in the early conceptual design phase of vehicle development where simplified models often are enough to analyze the structural behavior. An application of this approach is shown in [29]. Using topology optimization methods, the most efficient structures can be found in a relatively early stage of the vehicle design process. Topology optimization has proven to be very efficient, especially when used to improve existing designs [11].

The automotive body structure design problem has been considered in [24–26,68] by applying a sharing penalty vector method in order to define the product platform, that is, choose which components to share and design the product family with minimal individual variation from the ideal individually optimized designs. A multi-objective reliability-based approach has been followed in [44] for designing a piezoelectric brake.

An iterative response surface method has been used in [15] for the multidisciplinary optimization of a full vehicle to minimize mass while complying with crashworthiness and noise, vibration, and harshness (NVH) constraints.

The ATC process has also been applied to the optimal design of complete vehicle systems. Novel technologies, such as a hybrid-electric propulsion system, in-hub motors, and variable height suspensions, were considered in [67]. Emphasis was given to fuel economy, ride, and mobility characteristics.

In order to reduce the product complexity and cost while offering many variants, a widely implemented strategy is the product platform strategy, that is, sharing components and production processes across a platform of products.

Research topics in the automotive field include platform design process development [25,26] and optimum platform component selection [91].

Several decision-making models related to the acquisition process of vehicles that address a range of conflicting concerns, such as customer needs, lifecycle costs, new technology and design, budget allocation, competitive bidding, and the ability of contractors to meet market targets, are discussed in [13].

The advance in supercomputing has accelerated the development of simulation-based optimization [66]. The multidisciplinary design optimization for a car body structure under constraints of NVH and roof crush has been described in [111].

8.3.6 Integration of Vehicle Electronic Controls

Basic studies on active suspensions optimization and integration are given in [40,78,115]. As discussed previously, simple analytical formulae for the estimation of ride comfort and active safety of vehicles are derived and used to get the optimal controller gains that offer the best compromise among comfort, road holding, and packaging. In [36,37], this problem is reformulated by considering the vehicle body mass and the tire radial stiffness as stochastic parameters in order to derive the optimal trade-off solutions in a stochastic framework.

Research in several application areas has unveiled a coupling between the plant (design) and control optimization problems. Due to this coupling, optimizing a plant and control sequentially does not guarantee system optimality. The results have been demonstrated using a simple combined passive/active car suspension case study in [23].

A method to improve ride and handling of an automobile fitted with four active suspensions (sky hook dampers, active antiroll bars), active four-wheel steering, and traction control (controlled differential) has been proposed in [78]. The designer is asked to iterate between an innovation stage and a trade-off stage, in order to assess the effectiveness of candidate design issues for the synthesis of new control schemes.

In [94], a vehicle dynamics–integrated control algorithm designed by using a nonlinear optimization method is proposed for four wheel–distributed steering and four wheel–distributed traction/braking systems.

8.4 OPTIMIZATION OF THE DYNAMIC BEHAVIOR OF VEHICLES

A simple example is presented in the following to show how very general results can be obtained by applying correctly the optimization procedures introduced earlier.

The problem is to select the stiffness and damping of a passively suspended vehicle. The vehicle behavior, while running on a randomly profiled road, is described by means of the well-known quarter car vehicle model (see Figure 8.4a) and the objective is to minimize discomfort (the standard deviation of the acceleration of the body of the vehicle, $\sigma_{\ddot{x}_2}$), road holding (the standard deviation of the force acting between the road and the wheel, σ_{F_z}), and working space (the standard deviation of the relative displacement between the wheel and the vehicle body, $\sigma_{x_2-x_1}$).

By using the symbols shown in Figure 8.4a, we define

$$q = \frac{m_1}{m_2}, \quad K_x = k_2 \frac{(1+q)^2}{k_1 q}, \quad R_x = r_2 \sqrt{\frac{(1+q)^3}{k_1 m_2 q}} \tag{8.3}$$

to derive the analytical expressions of the pareto-optimal set into the (nondimensional) design variables domain (K_x, R_x) for the three combinations of two objective functions (boundaries of the pareto-optimal set for the problem with three objective functions, namely $\sigma_{\ddot{x}_2}$, σ_{F_z}, $\sigma_{x_2-x_1}$, see [37,39]). These are

$$\text{Optimal } (\sigma_{\ddot{x}_2}, \sigma_{F_z}) \rightarrow R_x = \sqrt{(1+q)K_x - qK_x^2} \quad \text{with } 0 \le K_x \le 1$$

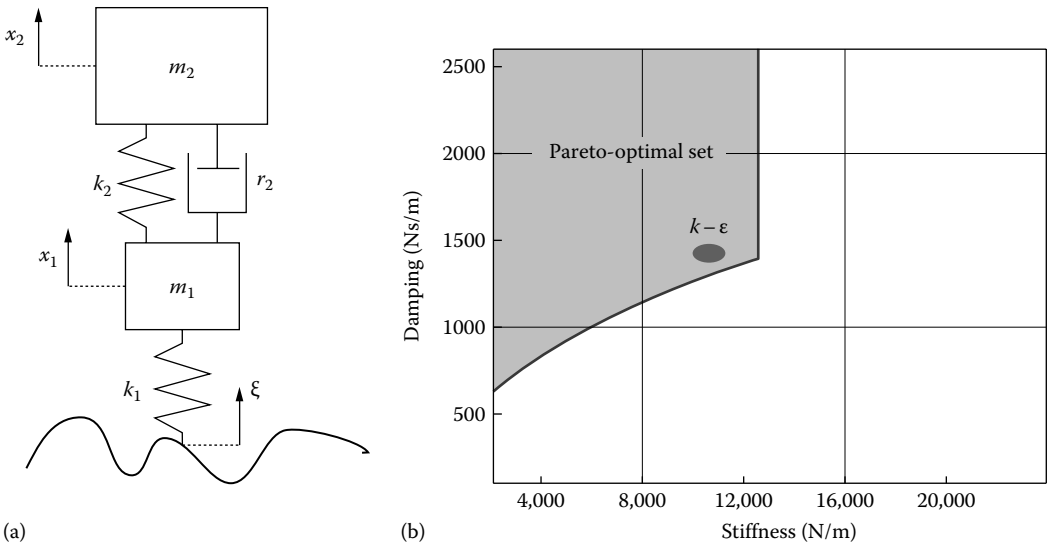


FIGURE 8.4 (a) Quarter car vehicle model, (b) Pareto-optimal set (filled area) and $k\varepsilon$-optimal solutions (see [72]) for the vehicle passive suspension optimization problem in the design variables space; suspension damping r_2 (Ns/m), suspension stiffness k_2 (N/m). Vehicle data are reported in [39]. k_1 (N/m) is the tire radial stiffness, m_1 (kg) is the unsprung mass, and m_2 (kg) is the sprung mass (for the quarter car model).

$$\text{Optimal } (\sigma_{\ddot{x}_2}, \sigma_{x_2-x_1}) \rightarrow R_x \geq 0, \quad K_x = 0 \tag{8.4}$$

$$\text{Optimal } (\sigma_{F_z}, \sigma_{x_2-x_1}) \rightarrow R_x \geq 1, \quad K_x = 1$$

The designer can get a direct insight into the suspension design problem by inspection of the expressions in Equation 8.4.

The pareto-optimal set into the design variable domain (suspension stiffness k_2 and suspension damping r_2) for the $\sigma_{\ddot{x}_2}$, σ_{F_z}, $\sigma_{x_2-x_1}$ problem is the filled area in Figure 8.4b (reference vehicle data are reported in [39]). The solutions belonging to the pareto-optimal set are equally desirable from the designer's point of view.

For the selection of the final suspension design solution, a $k\varepsilon$-optimality approach, extensively described in [72], has been followed. The result are the design solutions within the subset of the pareto-optimal set named $k\varepsilon$ (darker area) in Figure 8.4b. The result seems consistent with the solution selected by very skilled suspension specialists, after an expensive amount of trials.

In a stochastic formulation, the vehicle body mass m_2 and the tire radial stiffness k_1 can be considered as stochastic parameters due to the variety of possible loading conditions and to the variability of the tire pressure (m_2: CV = σ/μ = 0.10, k_1: CV = 0.10). The design variables to be optimized are the stiffness k_2 and the damping r_2 of the suspension considered as stochastic design variables due to variations with respect to their nominal value caused by production tolerances and/or wear (k_2: CV = 0.03, r_2: CV = 0.10). The computed pareto-optimal set is shown in Figure 8.5.

Figure 8.5 shows that the increase in robustness (higher reliability) implies a shift of the optimal solutions to higher values of the suspension stiffness k_2. It is shown that the pareto-optimal region computed by imposing a very high reliability (99.9999999%) is an extension of the deterministic one (see [37] for further details).

A similar approach can be used to optimize the handling behavior and active safety of road vehicles, as reported in [83]. The front and rear tyre cornering stiffness of a vehicle can be tuned in order to obtain a preferred compromise among the many conflicting performance indices that describe the handling behavior and active safety. The pareto-optimal set, shown in Figure 8.6, can be derived in analytical form.

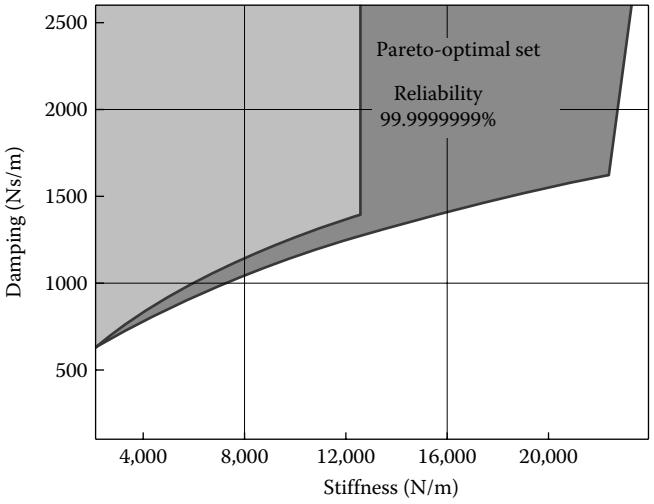


FIGURE 8.5 Pareto-optimal set (filled area) for the vehicle passive suspension optimization problem in the design variables space (suspension damping r_2 [Ns/m], suspension stiffness k_2 [N/m]), vehicle data reported in [39]. The deterministic solution (light color filled area) is reported together with the solution obtainable by applying a reliability-based optimization method (whole filled area).

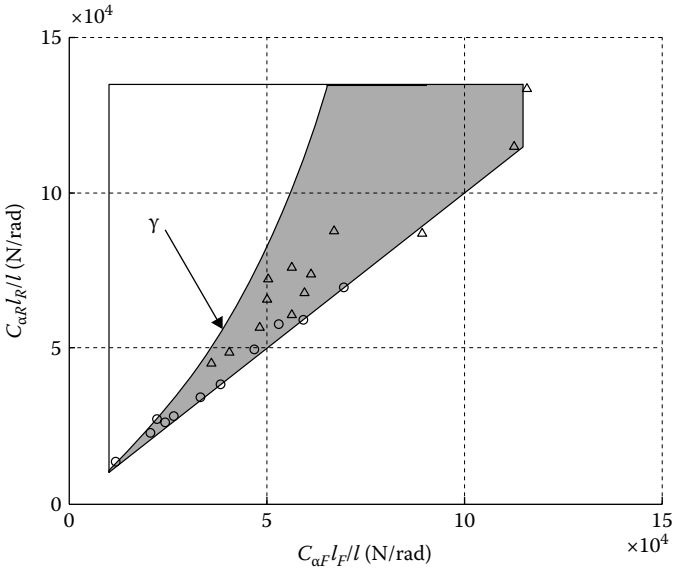


FIGURE 8.6 Pareto-optimal set (gray area) featuring the optimal handling performances of a road vehicle in the design variables domain $C_{\alpha F}/l_F$ and $C_{\alpha F}/l_F$. (From Mastinu, M. et al., *Optimal Design of Complex Mechanical Systems with Applications to Vehicle Engineering*, Springer Verlag, Berlin, Germany, 2006; Miano, C. et al., *ASME J. Mech. Design*, 126, 687, 2004.) Where $C_{\alpha F/R}$ is the front/rear cornering stiffness, $l_{F/R}$ is the respective distances from the center of gravity to the front/rear axles and l is the vehicle wheelbase. Experimental data denoted by ○ were derived from the literature; data denoted by Δ were available to the authors and measured by tire manufacturers. (Adapted from Miano, C. et al., *ASME J. Mech. Design*, 126, 687, 2004.)

8.5 CONCLUSION

A critical review of optimal design methods and their applications to solve vehicle system design problems has been presented. The correct optimization approach discards the "push the button and get the design" idea in favor of a realistic approach that recognizes the role of the designer as the leading force in the design process and the role of computers and software as indispensable tools. Successful optimization requires availability of appropriate analysis models and knowledge of the capabilities and limitations of the mathematical optimization techniques. Most of the modern software available for commercial use are capable of supporting such efforts effectively. However, an increased degree of familiarity is required before more complex studies are undertaken. The inclusion of uncertainty in the problem formulations is a relative recent development in the available optimization tools. Although solution robustness is desirable along with optimality, probabilistic optimal design formulations are substantially more computationally intensive than deterministic ones. Similarly, performing full pareto-optimality studies can be very expensive. The included (even still partial) list of references can be used to get a deeper insight into the subject and its potential practical implementations. These include problems in vehicle system dynamics, powertrain and internal combustion engine design, active safety and ride comfort, vehicle system design, lightweight structures, and controls.

REFERENCES

1. Ahn, Y.K. et al., Optimal design of an engine mount using an enhanced genetic algorithm with simplex method, *Vehicle System Dynamics*, 43(1), 57–81, 2005.
2. Alleyne, A. and Hedrick, J.K., Application of nonlinear control theory to electronically controlled suspensions, *Vehicle System Dynamics*, 22, 309–320, 1993.

3. Aminpour, M. et al., A framework for reliability-based MDO of aerospace systems (AIAA-2002-1476), *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Denver, CO, April 2002.
4. Ashby, M.F., Multi-objective optimization in material design and selection, *Acta Materialia*, 48, 359–369, 2000.
5. Baumal, A.E., McPhee, J.J., and Calamai, P.H., Application of genetic algorithms to the design optimization of an active vehicle suspension system, *Computer Methods in Applied Mechanics and Engineering*, 163, 87–94, 1998.
6. Belegundu, A.D. and Chandrupatla, T.R., *Optimization Concepts and Applications in Engineering*, Prentice Hall, Upper Saddle River, NJ, 1999.
7. Blouin, V., Fadel, G., Haque, I., Wagner, J., and Samuels, H., Continuously variable transmission, design for optimum vehicle performance by analytical target cascading, *International Journal of Heavy Vehicle Design—Special Automotive Research Center Edition*, 11(3/4), 327–348, 2004.
8. Caballero, R., Cerdá, E., Muñoz, M., Rey, L., and Stancu-Minasian, I.M., Efficient solution concepts and their relations in stochastic multiobjective programming, *Journal of Optimization Theory and Applications*, 110(1), 53–74, 2001.
9. Chakarvartula, S., Haque, I., and Fadel, G., A Monte-Carlo simulation approach to heavy vehicle design for good dynamic performance in multiple scenarios, *International Journal of Heavy Vehicle Systems* (*IJHVS*), 10(1/2), 112–143, 2003.
10. Chan, K.Y., Kokkolaras, M., Papalambros, P., Skerlos, S.J., and Mourelatos, Z., Propagation of uncertainty in optimal design of multilevel systems: Piston-ring/cylinder-liner case study, *Proceedings of SAE World Congress*, Detroit, MI, 2004, Paper No. 2004-01-1559.
11. Chen, C.J. and Usman, M., Design optimization for automobile applications, *International Journal of Vehicle Design*, 25, 126–141, 2001.
12. Cheng, F. and Li, D., Genetic algorithm development for multi-objective optimization of structures, *AIAA Journal*, 36(6), 1105–1112, 1998.
13. Cooper, A.B., Georgiopoulos, P., Michalek, J.J., and Papalambros, P.Y., A simulation-based vehicle design strategy for acquisition and requirements validation, *Proceedings of the SAE World Congress*, Detroit, MI, March 8–11, 2004.
14. Cooper, A.B., Kokkolaras, M., and Papalambros, P.Y., A dual-use enterprise context for vehicle design and technology valuation, *Proceedings of the Society of Automotive Engineers World Congress*, Detroit, MI, March 8–11, 2004, SAE Paper 2004-01-1588.
15. Craig, K., Stander, N., Dooge, D., and Varadappa, S., MDO of automotive vehicle for crashworthiness and NVH using response surface methods, AIAA-2002-5607, *Ninth AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, 2002.
16. Dahlberg, T., An optimized speed-controlled suspension of a 2-DOF vehicle travelling on a randomly profiled road, *Journal of Sound and Vibration*, 62, 541–546, 1979.
17. Davis, L., *The Handbook of Genetic Algorithms*, Van Nostrand Reingold, New York, 1991.
18. Deb, K., A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197, 2002.
19. De Weck, O., Multiobjective optimization: History and promise, Keynote Paper, *Third China–Japan–Korea Joint Symposium on Optimization of Structural and Mechanical Systems*, Kanazawa, Japan, October 30-November 2, 2004.
20. De Weck, O. and Jones, M.B., Isoperformance: Analysis and design of complex systems with desired outcomes, *Systems Engineering*, 9(1), 45–61, 2006.
21. Dong-Chan, L., Hyeun-Seok, C., and Chang-Soo, H., Design of automotive body structure using multicriteria optimization, *Structural and Multidisciplinary Optimization*, 32, 161–167, 2006 (doi: 10.1007/s00158-005-0577-2).
22. Ehrgott, M. and Wiecek, M.M., Multiobjective programming, in *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, and M. Ehrgott, eds., Springer, New York, 2005, pp. 667–722.
23. Fathyl, H., Papalambros, P., and Ulsoy, A., Integrated plant, observer, and controller optimization with application to combined passive/active suspension, *Proceedings of the 2003 ASME Mechanical Engineering Congress and Exposition*, Washington, DC, November 15–21, 2003.
24. Fellini, R., Papalambros, P.Y., and Weber, T., Application of a product platform design process to automotive powertrains, *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, 2000, Paper No. 4849.

25. Fellini, R. et al., A sensitivity-based commonality strategy for family products of mild variations, with application to automotive body structures, *Ninth AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, September 4–6, 2002, AIAA 2002-5610.
26. Fellini, R., Kokkolaras, M., and Papalambros, P., Quantitative platform selection in optimal design of product families, with application to automotive engine design, *Journal of Engineering Design*, 17(5), 429–446, October 2006.
27. Fenyes, P.A., Donndelinger, J., and Bourassa, J., A new system for multidisciplinary analysis and optimization of vehicle architectures, 2002, AIAA Paper, AIAA-2002-5509.
28. Filipi, Z. et al., Combined optimisation of design and power management of the hydraulic hybrid propulsion system for the 6 × 6 medium, *International Journal of Heavy Vehicle Systems*, 11(3/4), 372–402, 2004.
29. Fredricson, H., Property optimization of vehicle body structure elements using simplified models, *Proceedings of the WCSMO-5*, Lido di Jesolo, Italy, May 2003.
30. Fredricson, H., Johansen, T., Klarbring, A., and Petersson, J., Topology optimization of frame structures with flexible joints, *Journal of Structural and Multidisciplinary Optimization*, 25(3), 199–214, 2003.
31. Fujita, K. et al., Design optimization of a multilink suspension system for total vehicle handling and stability, *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St. Louis, MO, 1998.
32. Gobbi, M., Haque, I., Papalambros, P., and Mastinu, G., Optimization and integration of ground vehicle systems, *Vehicle System Dynamics*, 43(6–7), 437–453, June–July 2005.
33. Gobbi, M. and Mastinu, G., Global approximation: Performance comparison of different methods, with an application to road vehicle system engineering, in *Innovation in Vehicle Design and Development*, DE-Vol. 101, I. Haque et al., eds., ASME, New York, 1999, pp. 15–24.
34. Gobbi, M. and Mastinu, G., Stochastic multi-objective optimisation for the design of vehicle systems, *Proceedings of the WCSMO5 International Conference*, Lido di Jesolo, Italy, May 2003.
35. Gobbi, M. et al., Optimal and robust design of a road vehicle suspension system, *Vehicle System Dynamics Supplement*, 33, 3–22, 1999.
36. Gobbi, M., Levi, F., and Mastinu, G., Multi objective robust design of the suspension system of road vehicles, *Vehicle System Dynamics Supplement*, 41, 537–546, 2004.
37. Gobbi, M., Levi, F., and Mastinu, G., Multi-objective stochastic optimisation of the suspension system of road vehicles, *Journal of Sound and Vibration*, 298(4–5), 1055–1072, December 2006.
38. Gobbi, M. and Mastinu, G., On the optimal design of composite material tubular helical springs, *Meccanica*, 36(5), 525–553, 2001.
39. Gobbi, M. and Mastinu, G., Symbolic description and optimisation of the dynamic behaviour of vehicles running on rough road, *Journal of Sound and Vibration*, 245(3), 457–481, August 2001.
40. Gobbi, M. and Mastinu, G., Symbolical multi-objective optimisation of the dynamic behaviour of actively suspended road vehicles, *International Journal of Vehicle Design*, 28(1/2/3), 189–213, 2002.
41. Gobbi, M., Mastinu, G., and Caudano, M., Stochastic multi-objective optimisation of a gearbox synchroniser and selector mechanism, *Proceedings of the ASME IMECE 2003*, Vol. 2, Washington, DC, November 2003, ISBN 0-7918-4664-4.
42. Gobbi, M., Mastinu, G., D'Orazio, A., Caudano, M., and Faustini, G., On the optimisation of a double cone synchroniser for improved manual transmission shiftability, *Advanced Vehicle Technologies*, DE-Vol. 103, ASME, New York, 2002, ISBN 0-7918-1692-3.
43. Gobbi, M., Mastinu, G., and Doniselli, C., Optimising a car chassis, *Vehicle System Dynamics*, 32(2–3), 149–170, 1999.
44. Gobbi, M., Guarneri, P., and Mastinu, G., Optimal robust design optimization with application to a piezoelectric brake, SAE Paper 2008-01-2554. SAE Brake Colloquium & Exhibition, October 12–15, San Antonio, TX, 2008, ISBN 978-0-7680-2097-7.
45. Gobbi, M., Optimal and robust design of ground vehicle systems, *Proceedings of ESDA2006, 8th Biennial ASME Conference on Engineering Systems Design and Analysis*, Torino, Italy, July 4–7, 2006.
46. Goldberg, D., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
47. Goncalves, J. and Ambrosio, J., Optimization of vehicle suspension systems for improved comfort of road vehicles using flexible multibody dynamics, *Nonlinear Dynamics*, 34, 113–131, 2003.
48. Goncalves, J. and Ambrosio, J., Road vehicle modeling requirements for optimization of ride and handling, *Multibody System Dynamics*, 13, 3–23, 2005.
49. Grierson, D.E. and Hajela, P., eds., *Emergent Computing Methods in Engineering Design, Applications of Genetic Algorithms and Neural Networks*, Springer Verlag, Berlin, Germany, 1996.

50. Gu, L., Yang, R.J., Cho, C.H., Makowski, M., Faruque, M., and Li, Y., Optimization and robustness for crashworthiness, *International Journal of Vehicle Design*, 26(4), 348–360, 2001.
51. Hamza, K., Hossoy, I., Reyes-Luna, J.F., and Papalambros, P.Y., Combined maximization of interior comfort and frontal crashworthiness in preliminary vehicle design, *International Journal of Vehicle Design*, 35(3), 167–185, 2004.
52. Hedrick, J.K., Introduction to active suspensions, Carl-Cranz-Academy Course Notes on "Control of Vehicle Ride and Handling", Wessling-Oberpfaffenhofen, July 1990.
53. Holland, J., *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, MI, 1975.
54. Hrovat, D., Applications of optimal control to advanced automotive suspension design, *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, 115, June 1993.
55. http://frontier.enginsoft.it/ (May 12, 2012).
56. http://www.altair.com/ (May 12, 2012).
57. http://www.ichrome.eu/ (May 12, 2012).
58. http://www.mathworks.com/ (May 12, 2012).
59. http://www.mscsoftware.com/ (May 12, 2012).
60. http://www.noesis.be/ (May 12, 2012).
61. http://www.vrand.com/ (May 12, 2012).
62. http://www.3ds.com/ (May 12, 2012).
63. Karnopp, D. and Trikha, A., Comparative study of optimization techniques for shock and vibration isolation, *Transaction of the ASME, Journal of Engineering for Industry*, 91, 4, 1969.
64. Kim, H., Rideout, D., Papalambros, P., and Stein, J., Analytical target cascading in automotive vehicle design, *Journal of Mechanical Design*, 125, 481, September 2003.
65. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., Optimization by simulated annealing, *Science*, 220(4598), 1983, 13.
66. Kodiyalam, S., Yang, R.J., Gu, L., and Tho, C., Large-scale, multidisciplinary optimization of a vehicle system in a scalable, high performance computing environment, *Proceedings of DETC2001 ASME Design Engineering Technical Conferences*, Pittsburg, PA, September 9–12, 2001.
67. Kokkolaras, M. et al., Simulation-based optimal design of heavy trucks by model-based decomposition: An extensive analytical target cascading case study, *International Journal of Heavy Vehicle Systems*, 11(3/4), 403–433, 2004.
68. Kokkolaras, M., Fellini, R., Kim, H.M., Michelena, N.F., and Papalambros, P.Y., Extension of the target cascading formulation to the design of product families, *Journal of Structural and Multidisciplinary Optimization*, 24(4), 293–301, 2002.
69. Kokkolaras, M., Mourelatos, Z.P., and Papalambros, P.Y., Design optimization of hierarchically decomposed multilevel system under uncertainty, *Proceedings of the ASME 2004 Design Engineering Technical Conferences*, Salt Lake City, UT, September 28–October 2, 2004, DETC2004/DAC-57357.
70. Kokkolaras, M., Mourelatos, Z.P., and Papalambros, P.Y., Design optimization of hierarchically decomposed multilevel systems under uncertainty, *Proceedings of 2004 ASME Design Engineering Technical Conferences: 30th Design Automation Conference*, Salt Lake City, UT, 2004, Paper No. DETC2004/DAC57357.
71. Levi, F., Gobbi, M., and Mastinu, G., An application of multi-objective stochastic optimisation to structural design, *Structural and Multidisciplinary Optimization*, Springer-Verlag, Berlin, Germany, No. 10, 2004, DOI: 10.1007/s00158-004-0456-2.
72. Gobbi, M., A k, k-ε optimality selection based multi objective genetic algorithm with applications to vehicle engineering Optimization and Engineering, January 2012.
73. Li, Z., Georgiopoulos, P., Papalambros, P.Y., Filipi, Z., Wu, G., and Yang, X., Model based analysis of performance-cost tradeoffs for engine manifold surface finishing, *Proceedings of the Society of Automotive Engineers World Congress*, Detroit, MI, March 8–11, 2004, SAE Paper 2004-01-1561.
74. Li, Z., Kokkolaras, M., Jung, D., Papalambros, P.Y., and Assanis, D.N., An optimization study of manufacturing variation effects on diesel injector design with emphasis on emissions, *Proceedings of the SAE World Congress*, Dearborn, MI, March 8–11, 2004, SAE Paper 2004-01-1560.
75. Longo, R., Anedda, G., and Testa, M., Integrated multibody structural optimization for car suspension design, *Proceedings of the TCN CAE 2003 International Conference*, Cagliari, Italy, 2003.
76. Mastinu, G., Automotive suspension design by multiple objective programming, *Proceedings of the International Symposium on "Advanced Vehicle Control" (AVEC'94)*, JSAE, Tsukuba, Japan, 1994, pp. 73–78.

77. Mastinu, G. and Gobbi, M., Advances in the optimal design of mechanical systems, Course coordinated by CISM (International Centre for Mechanical Sciences), Birla Science Centre, Hyderabad, India, www.europeindia.org, 1999.
78. Mastinu, G., Integrated controls and interactive multiple objective programming for the improvement of ride and handling of road vehicles, in *Smart Vehicles' Swets & Zeitlinger*, J.P. Pawelussen and H. Pacejka, eds., Lisse, the Netherlands, 1995.
79. Mastinu, M., Gobbi, M., and Miano, C., *Optimal Design of Complex Mechanical Systems with Applications to Vehicle Engineering*, Springer Verlag, Berlin, Germany, 2006, ISBN 3-540-34354-7.
80. Metropolis, N., Rosenbluth, A., Rosenbluth, R., Teller, A., and Teller, E., Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, 21, 1087–1092, 1953.
81. Matusov, J.B., *Multicriteria Optimisation and Engineering*, Chapman & Hall, New York, 1995.
82. Miano, C., Gobbi, M., and Mastinu, G., A tutorial on present and future applications of global approximation issues with application to vehicle design problems, in *Advanced Vehicle Technologies*, DE-Vol. 112, ASME, New York, 2001.
83. Miano, C., Gobbi, M., and Mastinu, G., Multi-objective optimization of the handling performances of a road vehicle: A fundamental study on tire selection, *ASME Journal of Mechanical Design*, 126, 687, July 2004.
84. Miano, C., Gobbi, M., Mastinu, G., and Cesarini, R., On the integrated design of the tyre-suspension system of a racing car, in *Advanced Vehicle Technologies*, DE-Vol. 106, I. Haque et al., eds., ASME Publication, New York, 2000.
85. Miano, C., Optimization methods of complex mechanical system with particular reference to ground vehicles (In Italian). PhD thesis, Politecnico di Milano, Como, Italy, 2003.
86. Michelena, N.F. and Papalambros, P.Y., A hypergraph framework for optimal model-based decomposition of design problems, *Journal of Computational Optimization and Applications*, 2, 8, 1997.
87. Michelena, N.F., Park, H., and Papalambros, P.Y., Convergence properties of analytical target cascading, *AIAA Journal*, 41(5), 897–905, May 2003.
88. Miettinen, K., *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, MA, 1999.
89. Mitschke, M., *Dynamik der Kraftfahrzeuge*, Springer Verlag, Berlin, Germany, 1990.
90. Nakhaie Jazar, G., Alkhatib, R., and Golnaraghi, M.F., Root mean square optimization criterion for vibration behaviour of linear quarter car using analytical methods, *Vehicle System Dynamics*, 44(6), 477–512, 2006.
91. Nelson, S.A., II, Parkinson, M.B., and Papalambros, P.Y., Multicriteria optimization in product platform design, *Journal of Mechanical Design*, 123(2), 199–204, 2001.
92. Ng, W.Y., *Interactive Multi-Objective Programming as a Framework for Computer Aided Control Systems Design*, Springer Verlag, Berlin, Germany, 1989.
93. Nishiwaki, S., Nishigaki, H., Tsurumi, Y., Kojima, Y., and Kikuchi, N., First order analysis—New CAE tool for automotive body designers, 2001, SAE Paper 2001-01-0768.
94. Ono, E., Hattori, Y., Muragishi, Y., and Koibuchi, K., Vehicle dynamics integrated control for four-wheel-distributed steering and four-wheel-distributed traction/braking systems, *Vehicle System Dynamics*, 44(2), 139–151, 2006.
95. Papalambros, P.P. and Wilde, D.J., *Principles of Optimal Design: Modeling and Computation*, Cambridge University Press, New York, 2000.
96. Pardalos, P.M., Resende, M., and Pardalos, P., *Handbook of Applied Optimization*, Oxford University Press, Oxford, U.K., 2002.
97. Ruzika, S. and Wiecek, M.M., A survey of approximation methods in multiobjective programming, *Journal of Optimization Theory and Applications*, 126(3), 473–501, 2005.
98. Scarlat, G., Haque, I., Fadel, G., and Schuller, J., A modified Monte-Carlo approach to simulation-based vehicle design with multiple objectives and multiple scenarios, *Proceedings of the SAE World Congress*, Detroit, MI, 2002.
99. Schäffer, J., Some experiments in machine learning using vector evaluated genetic algorithms, PhD thesis, Vanderbilt University, Nashville, TN, 1984.
100. Schuller, J., Haque, I., and Fadel, G., Methods for simulation-based optimization of vehicle handling behavior, in *Advanced Vehicle Technologies*, I. Haque and M. El-Gindy, eds., ASME, New York, November 2000.
101. Schuller, J., Haque, I., and Eckel, M., An approach for optimisation of vehicle handling behaviour in simulation, *Dynamics of Vehicles on Roads and Tracks, Supplement to Vehicle System Dynamics*, 37, 24–37, August 2001.
102. Sharp, R.S. and Crolla, D.A., Road vehicle suspension system design—A review, *Vehicle System Dynamics*, 16, 167–192, 1987.

103. Sinha, K., Reliability-based multi-objective optimization methodology for engineering systems, *Proceedings of ESDA2006, 8th Biennial ASME Conference on Engineering Systems Design and Analysis*, Torino, Italy, July 4–7, 2006.
104. Siddal, J.N., A new approach to probability in engineering design and optimization, *ASME Journal of Mechanisms, Transmissions, and Automation in Design*, 106, 5–10, 1984.
105. Srinivas, N. and Deb, K., Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation*, 2(3), 221–248, 1994.
106. Suh, M.-W., Le, J.-H., Cho, K.-Y., and Kim, S.-I., Section property method and section shape method for the optimum design of a vehicle body structure, *International Journal of Vehicle Design*, 30, 115–134, 2002.
107. Suppapitnarm, A., Seffen, K.A., Parks, G.T., Clarkson, P.J., and Liu, J.S., Design by multiobjective optimization using simulated annealing, *Proceedings of the International Conference on Engineering Design ICED*, Munich, Germany, 1999.
108. Tangella, C., Simulated annealing as an optimization tool for multi-criteria, multi-scenario optimization, MS thesis, Clemson University, Clemson, SC, May 2003.
109. Tezuka, D., *Uniform Random Numbers: Theory and Practice*, Kluwer Academic Publishers, Boston, MA, 1995.
110. Thompson, A., An active suspension with optimal linear state feedback, *Vehicle System Dynamics*, 5(4), 187–203, 1976.
111. Thompson, A., Suspension design for optimum road-holding, 1983, SAE Paper 830663.
112. Yang, R.J., Gu, L., Tho, C.H., and Sobieski, J., Multidisciplinary design optimization of a full vehicle with high performance computing, *42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Seattle, WA, 2001.
113. Youn, B.D. et al., Techniques for estimating uncertainty propagation in probabilistic design of multilevel systems, *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, NY (AIAA 2004-4470), August 30–September 1, 2004.
114. Youn, B.D. and Choi, K.K., Selecting probabilistic approaches for reliability-based design optimization, *AIAA*, 42(1), 124–131, 2004.
115. Zaremba, A., Hampo, R., and Hrovat, D., Optimal active suspension design using constrained optimization, *Journal of Sound and Vibration*, 207(3), 351–364, 1997.

9 Fatigue and Structural Durability of Automotive Components

Thomas Bruder, Holger Hanselka, Rüdiger Heim, Heinz Kaufmann, Michael Kieninger, Jürgen Nuffer, and Cetin M. Sonsino

CONTENTS

9.1 Structural Durability 264
9.1.1 Influencing Parameters 266
9.1.1.1 Material, Geometry, and Manufacturing 266
9.1.1.2 Loading 268
9.1.2 Cyclic Data 269
9.1.2.1 Presentation of Cyclic Data 269
9.1.2.2 Strain Woehler Curves 270
9.1.2.3 Significance of Tests with VAs 272
9.1.2.4 Size Effect and Transferability 272
9.1.3 Component Design/Assessment 273
9.1.3.1 Initial Design/Assessment Data 273
9.1.3.2 Design Concepts 274
9.1.3.3 Strength Hypotheses 275
9.1.3.4 Estimation of Fatigue Life 276
9.1.4 Endurable Loading 277
9.1.4.1 Fracture Behavior Subject to Impact Loading 277
9.1.4.2 Structural Yield Strength 278
9.1.4.3 Choice of Relevant Design Data 279
9.1.4.4 Safety Considerations—Endurable and Allowable (Permissible) Stresses 279
9.1.4.5 Experimental Proof Tests 280
9.1.4.6 Notations 281
9.2 Load Data Acquisition 282
9.2.1 Acquisition of Load Data 282
9.2.2 Data Acquisition of Stresses 284
9.3 Derivation and Generation of Loads for Evaluation and Testing 285
9.3.1 Fundamentals and Influences on Fatigue Life 285
9.3.1.1 Mission Analysis and Usage Monitoring 287
9.3.1.2 Loading Sequence and Load Amplitude Variation 287
9.3.1.3 Influence of Overloads on Fatigue 290
9.3.2 Extrapolation of Measurements 290
9.3.3 Shortening of Load-Time Sequences 290
9.3.3.1 Omission 291
9.3.3.2 Truncation 293
9.3.3.3 Elastomeric Components 293

9.3.4 Synthesis of Load Data 293
9.3.4.1 Combination of Loading Segments 293
9.3.4.2 Frequency-Based Analyses 297
9.3.5 Standardized Load Sequence CARLOS TC 297
9.4 Accelerated Life Testing by Laboratory Road Simulation 297
9.4.1 Road Testing and Road Simulation 299
9.4.2 Definition of System to Be Tested 300
9.4.2.1 Spindle-Coupled Road Simulators 300
9.4.2.2 Wheel and Hub Road Simulators 301
9.4.3 Load File Development 302
9.4.4 Evaluation of Test Results 302
References 303

This chapter aims to give a short overview on structural durability as understood and practised at present. It is subdivided into the following four sections:

Section 9.1 gives a short overview of structural durability. It explains the influencing parameters material, geometry, manufacturing, and loading, defines stress and strain Woehler curves and Gassner curves and the transferability of fatigue data. For design purposes, different concepts are discussed. This section shows that structural durability cannot be handled as a material characteristic only, but must always be seen also under the aspect of a component and load-related material behavior.

Section 9.2 deals with the acquisition of loads. Measurements conducted under service conditions to determine damage parameters relevant to structures form the basis of experimental and numerical simulations, strength evaluations, and standardization. Special load sensors first enabled data acquisition technology to measure internal loads and efficient acquisition devices and sensors are now available. Nevertheless, the challenge of recording measurements under service conditions is always associated with a certain degree of effort. A systematic application of measurement acquisition of loads and stresses and strains substantially increases the reliability of components and assemblies being developed. Even during the predevelopment stages, the use of previously measured loads can be utilized and in the development process itself, the load assumptions can be finely adjusted and verified. Thus, load data acquisition is to be seen as an indispensable instrument in the development of reliable components and systems.

Realistic assumptions about loads are essential for a sound analysis of structural durability. Section 9.3 gives a brief overview on how to derive and generate the loading for numerical evaluation as well as physical testing purposes.

In Section 9.4, it will be shown that experimental proof is an essential measure for reliability, robustness, and quality. Since many vehicle components are relevant for operational safety, these parts must not fail and cause an accident. This can be ensured by laboratory road simulation. Hence, service load simulation and evaluation becomes an instrument to actively guide the product development process, rather than to release newly developed parts only. If both loading and assembly interactions are incorporated in a realistic manner, cost and time are saved due to the elimination of specialized on-track measurements.

Typically, multiaxial test rig hardware is used together with digital controllers for forces and motion. This leads to accelerated life testing on a subsystem level, for example, chassis and suspension or unsprung rotating components such as wheels and hubs. The load files are derived from road measurements or using scalable standardized information on the loading.

9.1 STRUCTURAL DURABILITY

Structural durability (service strength) is a generic term for a spectrum of stresses (loadings) that a component must endure during its required life without failure. Basic parameters are the material, design (component geometry), manufacturing process (e.g., casting, forging, welding, or surface treatment),

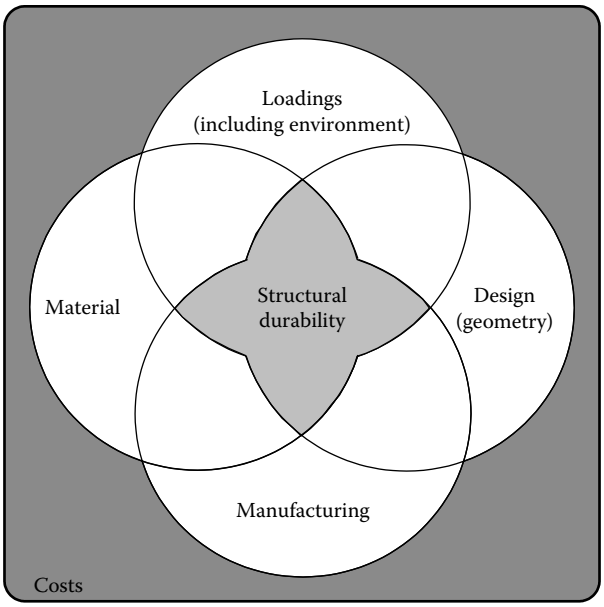


FIGURE 9.1 Parameters influencing the structural durability of components.

and loading (mechanical as well as superimposed environmental influences) [1–6]; the costs are ultimately responsible for the fact that products come on to the market optimized rather than optimal (Figure 9.1).

Despite many years of research, the release of reliable components on the basis of a purely numerical proof of structural durability is not possible. Not only the shortened development time but also the complexity of components and, in some cases, integration of, for example, smart structures, demand even more numerical analysis in order to be able to competently accompany future development processes and to meet requirements for product reliability. The methodological development, used to broaden the basis for design and optimization, extends to parameters such as manufacturing methods, surface state, size effect, and strength hypotheses. The transferability of fatigue data to components is closely associated with both local microstructural properties and failure mechanisms and requires the support of metallographic and scanning electron microscopic investigations. The present understanding of structural durability distinguishes itself from the original definition introduced by Ernst Gassner in 1939, which equated structural durability with fatigue behavior under variable amplitude (VA) loading. The definition today comprises

- Special loadings (overloads, which, for example, approach the component's structural, yield strength, for example, chassis components whilst slowly driving over an obstacle), bulging, or buckling of structures and misuse through dynamic loading (high speed and energy impact loading, for example, driving at speed over a pothole), where brittle fracture without deformation due to incorrect design and/or material choice must be excluded; that is, ductile fracture could be admissible
- Creep loading, for example, turbine blades subject to high temperature and centrifugal force
- Wear between frictional parts, for example, gear wheels
- Cyclic (not dynamic) loading (fatigue strength) with constant amplitudes (CAs) or VAs with superimposed assembly or loading-dependent mean values can lead to material fatigue, crack initiation, and finally to failure of components. The strength behavior under CA is

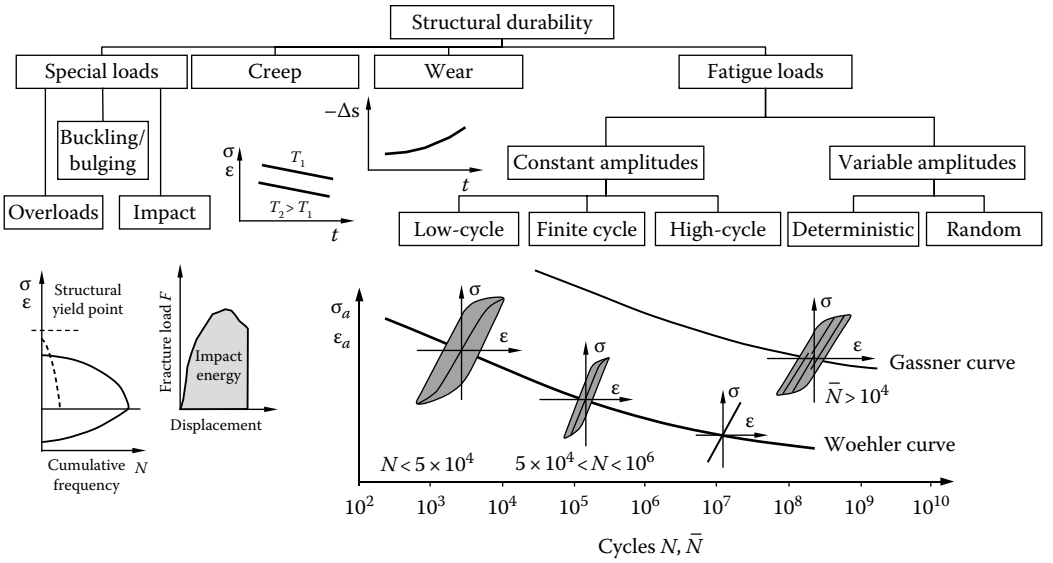


FIGURE 9.2 Partition of structural durability.

described by the Woehler curve (S–N curve: endurable stress/strain amplitudes under CA loading; in this section, denoted as "Woehler" curve in order to distinguish it from the endurable stress/strain amplitude loading, i.e., the Gassner curve) and under VA by the Gassner curve (Figure 9.2)

The Woehler curve is divided into three regions: low-cycle fatigue (elastoplastic deformations with a larger plastic part, defined up to 5×10^4 cycles), finite fatigue life region (elastoplastic deformations, endurance between 5×10^4 and 1×10^6 cycles), and the so-called "endurance limit" (macroscopical elastic deformations with endurances greater than 1×10^6 cycles). In this region, one must anticipate a continuous decrease in fatigue strength depending on material, manufacturing method, and environment. That is, an "endurance limit" does not exist [7,8]. The Gassner curve can be extended between 10^4 and more cycles up to 10^{10} while the VA loading can occur deterministically or stochastically. Typical components lying in the low-cycle fatigue region are reactor pressure vessels; those in the finite fatigue life region are vehicle door hinges and those in the so-called endurance limit region are gear wheels, crank-shafts, or con-rods. Typical components that are designed for variable service loads are chassis parts, for example, wheels, steering rod, front and rear axles, stub axles, braking systems, etc. The types of loading described here (special events, creep, wear, and cyclic loading) seldom occur individually in service; they can by all means occur successively or together such that their fatigue life defining interaction must be considered. However, loadings or stresses with only CAs, be it in the region of low-cycle fatigue, finite fatigue life, or so-called endurance limit, occur very seldom. In service, a composition of amplitudes of different magnitudes mainly exists, which is described by the term spectrum (cumulative frequency distribution of amplitudes). A knowledge of this spectrum is a prerequisite for determining the Gassner curves.

9.1.1 Influencing Parameters

9.1.1.1 Material, Geometry, and Manufacturing

In general, an increase in the values of ultimate tensile or yield strength for metallic materials is accompanied by an increase in the fatigue strength. However, this applies only to unnotched or lightly notched component regions, since the notches, dependent on design, significantly, influence

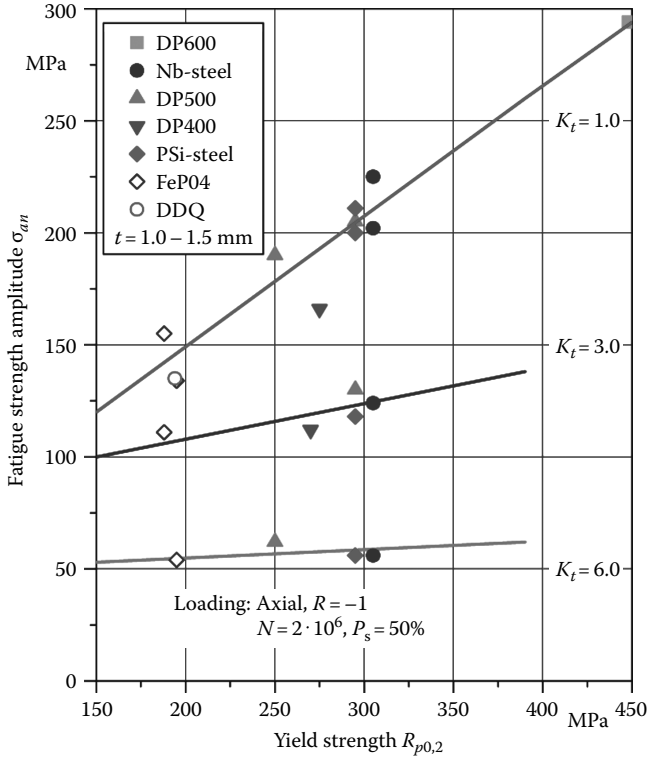


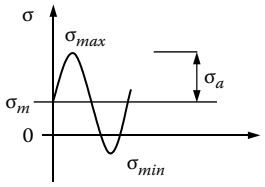
FIGURE 9.3 Fatigue strength of car body steels subjected to the notch factor K_t.

the fatigue strength. Notches with increased stress concentration (e.g., shoulders, fillets, bore holes, etc. with small radii) lower the external tolerable load. For notches, with high stress concentrations ($K_t > 2.5$), the fatigue strength cannot be raised by choosing a high strength material (Figure 9.3).

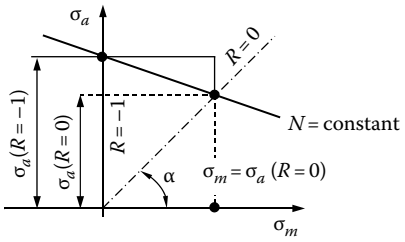
Therefore, the material's higher strength can only be exploited when stress concentrations are explicitly reduced by design, in most cases by incorporating larger radii. In cases where it is not possible to reduce stress concentrations, mechanical or thermomechanical surface treatment operations can be added to the manufacturing process. However, materials with higher tensile strength or yield strength also experience greater enhancement in fatigue strength than materials with lower strengths [9]. This enhancement is due to the material's ability to establish higher and stable compressive residual stresses. During mass production, effective options for enhancing fatigue strength are implemented, for example, surface rolling or induction hardening. According to the material, geometry, and loading (e.g., fully reversed, pulsating, CAs or VAs), enhancement of the endurable stress amplitudes up to a factor of 3 may be reached by these options under CA loading. Under VA loading, depending on the load level, the observed increase can be reduced due to the diminishing of compressive residual stresses by local plastic deformations [5]. This enhancement is dominated by the process-dependent compressive residual stresses introduced into the surface layer. The fatigue strength can be frequently enhanced more effectively using these measures than by design features. Basically, all surface layer states affecting fatigue strength are to be assessed from the perspective of "manufacturing induced" residual stresses, for example, milled, turned, and ground. A further important parameter is the mean stress sensitivity M, which serves to assess the influence of assembly and loading-dependent mean stresses on the endurable fatigue strength (Figures 9.4 and 9.5).

Mean stress sensitivity M and stress ratio R:

$$M = \frac{\sigma_a(R=-1)}{\sigma_a(R=0)} - 1 \qquad R = \frac{\sigma_{min}}{\sigma_{max}} = \frac{\sigma_m - \sigma_a}{\sigma_m + \sigma_a}$$



Mean stress and amplitude in the Haigh-diagram:



$$\sigma_a = \sigma_m \cdot \frac{1-R}{1+R} = \sigma_{max} \cdot \frac{1-R}{2}$$

FIGURE 9.4 Parameter of the Haigh diagram.

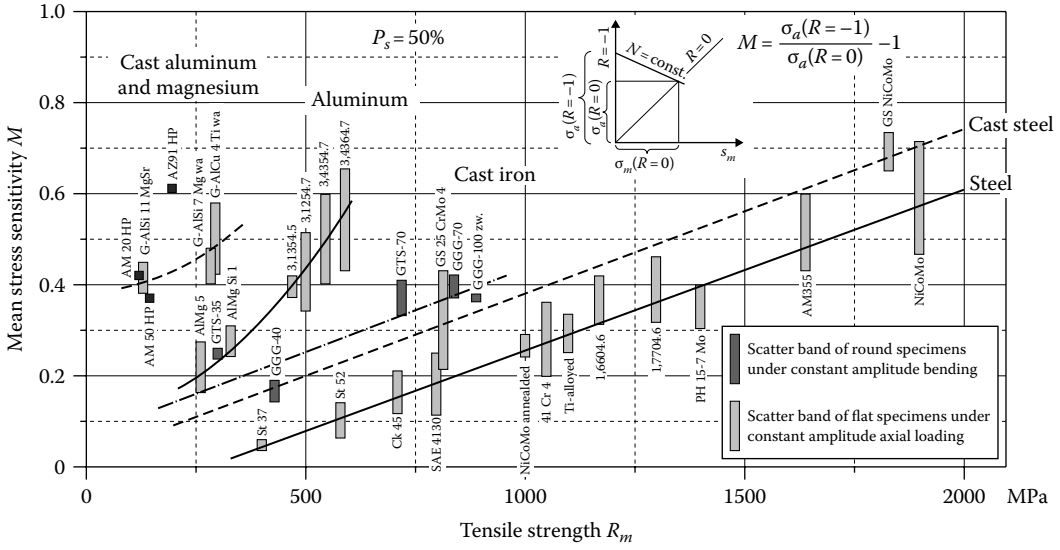


FIGURE 9.5 Mean stress sensitivity of some metallic materials.

M describes the ratio of the endurable stress amplitude σ_a for a material subject to fully reversed loads ($R = -1$) and the pulsating loads ($R = 0$). For all metallic materials, tensile mean stresses σ_m diminish the endurable stress amplitude σ_a, compressive mean stresses elevate σ_a. If stress ratios occurring in service deviate from fully reversed ($R = -1$) or pulsating loadings ($R = 0$), the fatigue strength amplitudes belonging to the respective mean stress can be approximately derived from the Haigh diagram by interpolation or extrapolation (Figure 9.4). However, for data lying outside the range $-1 \leq R \leq 0$, experimental verification is recommended.

9.1.1.2 Loading

In addition to the applied mechanical loadings (CAs or VAs, uniaxial or multiaxial loading), environmental influences such as temperature or corrosive media on the fatigue strength must be considered. Oxidation and creep effects can also occur during long holding periods, for example, stationary operations at high temperatures. Depending on their chloride ion concentration, solutions

containing chlorides very severely degrade the fatigue life (or the corrosion fatigue strength), since the Woehler curve's knee point is shifted to higher numbers of cycles or disappears completely. The loss in fatigue strength produced by a corrosive environmental medium is generally lower for VAs (service load sequences) than for CAs [10]. The extent to which coatings, for example, such as thermal spraying, galvanizing, etc., can be effective against corrosion fatigue must be investigated for individual cases. Examples of other media, which must also be considered during the design assessment, are fuel and brake fluid. A further type of corrosion that takes place under cyclic loading is fretting fatigue. This occurs by the relative movement between mating materials. Fretting fatigue can be very effectively reduced by shot peening and surface rolling (surface work hardening with induced compressive residual stresses) [11]. Surface treatments for steels such as nitriding, induction or case hardening (surface hardening with induced compressive residual stresses) would be expected to increase fatigue strength.

9.1.2 Cyclic Data

9.1.2.1 Presentation of Cyclic Data

The Woehler curve is a double logarithmic plot of loading or stress or strain amplitude (often stress) against the fatigue life (number of cycles, endurance), where a failure criterion (total rupture, a defined crack size, a reduction in stiffness, etc.) for the fatigue life must be defined. The significant influencing parameters are the material, the theoretical notch factor K_t (provided it can be defined for the section to be assessed), the loading or stress ratio $R = X_{min}/X_{max}$ (information about the mean value on which the amplitude is superimposed), type of loading (axial, bending, torsion), and the environment (temperature, medium). Loads are external parameters such as force, moment, pressure, or parameters resulting locally in the component strain, stress, or deformation (Figure 9.6).

The parameters for describing a Woehler curve are the slope k in the finite fatigue life region, the numbers of cycles at the knee point N_k, stress (or strain) amplitude at the knee point, for example, σ_{ak}, and the slope k' following the knee point; a probability of survival or failure P_s or P_f respectively, is assigned to these parameters. Both the height and position of the knee point together with the slopes are dependent on the influencing parameters. The scatter of the results T (in loading or life) is dependent on the material and manufacturing process, and is necessary for a safety

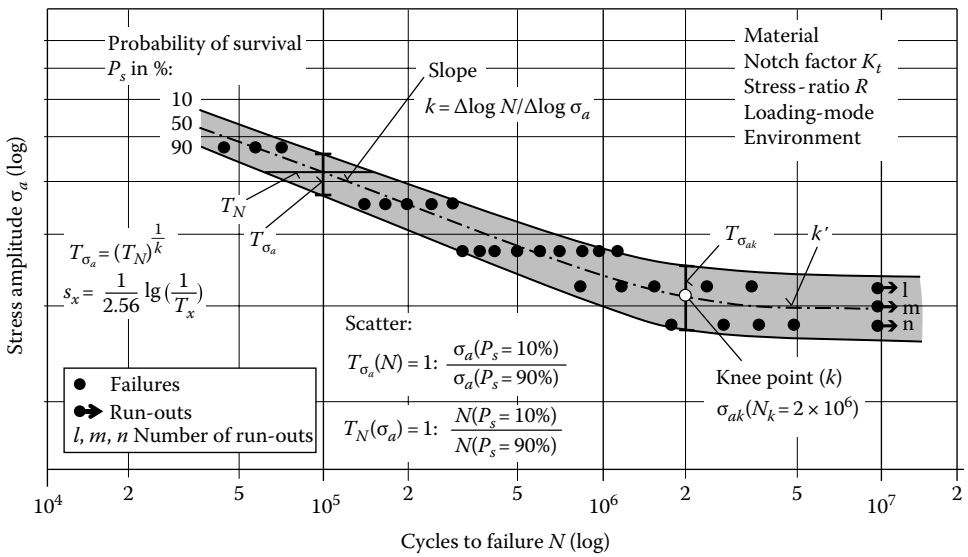


FIGURE 9.6 Parameters to determine and describe the Woehler curve.

consideration, whose computation frequently uses the Gaussian log-normal distribution. Under this assumption, the standard deviation can be computed from the scatter T, which is defined as the ratio between the probability of survivals $P_s = 10\%$ and 90%, $T_X = 1{:}X(P_s = 10\%)/X(P_s = 90\%)$ (Figure 9.6). The scatters T_N and $T_{\sigma a}$ in the region of finite fatigue life are coupled through the slope k. The scatter bands become wider toward the knee point with increasing fatigue life. The designation of scatter of fatigue life N for the region $N > N_k$ makes no sense because of the Woehler curve's shallow course following the knee point and may result in $T_N = \infty$. Therefore, for this region only, the scatter T of the loading or stress amplitude is determined. This does not become greater than the scatter in the transition region from the steep to the shallow sloped regions of the Woehler curve. That is, the scatter determined in the region of finite fatigue life just before the knee point can be considered to remain constant after the knee point. This assumption is expedient, since the stair case method [12] and its modifications, which are used to determine the endurable load, strain, or stress amplitudes for a fixed number of cycles at or after the knee point, provide only a reliable mean value but no reliable scatter. The aforementioned parameters are correspondingly valid for the determination and description of the Gassner curve (identified by "—" and denotes "the peak value of loading"), although the scatter in the Gassner curve can be lower than that in the Woehler curve. According to the shape of the spectrum, the slopes of Gassner curves can be often shallower than those of Woehler curves ($\overline{k} \geq k$). Generally, both can be referred to as fatigue life curves. Only by plotting on a double logarithmic scale does the representation give a straight line with a slope k, k', or $\overline{k}$. As a rule, the amplitudes are plotted on the loading or stress axis; a possible superimposed mean value emerges from the R-value. On determining Gassner curves, a subsequence, for example, as peak value sequence with a defined sequence length, must exist [13,14]. A load spectrum is the sum of the endured subsequences. The tests are carried out at different load levels whereby the number of subsequence repetitions and the magnitude of the amplitudes are level-dependent. A higher and a lower level merely define a linear increase and decrease, respectively, of all amplitudes and their associated mean values (Figure 9.7).

In order to represent service loading conditions, a subsequence must be repeated at least 5–10 times up to failure in order to provide a service-like mixing. The results of a Gassner test are plotted (at least in German-speaking countries) against the amplitude's peak values of each applied subsequence and resulting number of cycles to failure. Considering the R-value, this enables a comparison with the material's yield strength or component's structural yield strength. Furthermore, it is discernable how much the peak value exceeds the knee point of the Woehler curve [15,16] (the so-called endurance limit). This gives information about the lightweight design possibilities [13,14,17]. The precise position of the Gassner curve, as a function of the used sequence, can be theoretically computed from the damage accumulation [5]. However, it is only reliably determined by tests.

9.1.2.2 Strain Woehler Curves

To assess the material behavior for loading conditions exceeding the yield stress, it is necessary to know both the tensile and cyclic stress–strain curves together with the crack initiation Woehler curve under strain-controlled conditions [18] for assessment of the elastoplastic material behavior. The basis of dimensioning components, according to the local strain concept, is the simulation of cyclic behavior at notch roots of components using the results of strain-controlled tests on unnotched specimens. Owing to stress gradients in highly strained/stressed regions of components, the local deformations behavior is strain-controlled, provided that the structural yield point is not exceeded. Strain Woehler curves are the basis for component design according to the local strain approach. (In the macroscopically elastic region, strains and stresses are convertible to each other by the E-modulus, so that in this region local strain and stress concepts are identical.) The coefficients and exponents according to Coffin–Manson–Basquin, with which the strain Woehler curve is described, and those according to Ramberg–Osgood for the monotonic and cyclic stress–strain curve are used as input data for commercial dimensioning software [13,19]. By contrasting the monotonic with the cyclic stress–strain curve, it can be seen whether by repeated elastoplastic

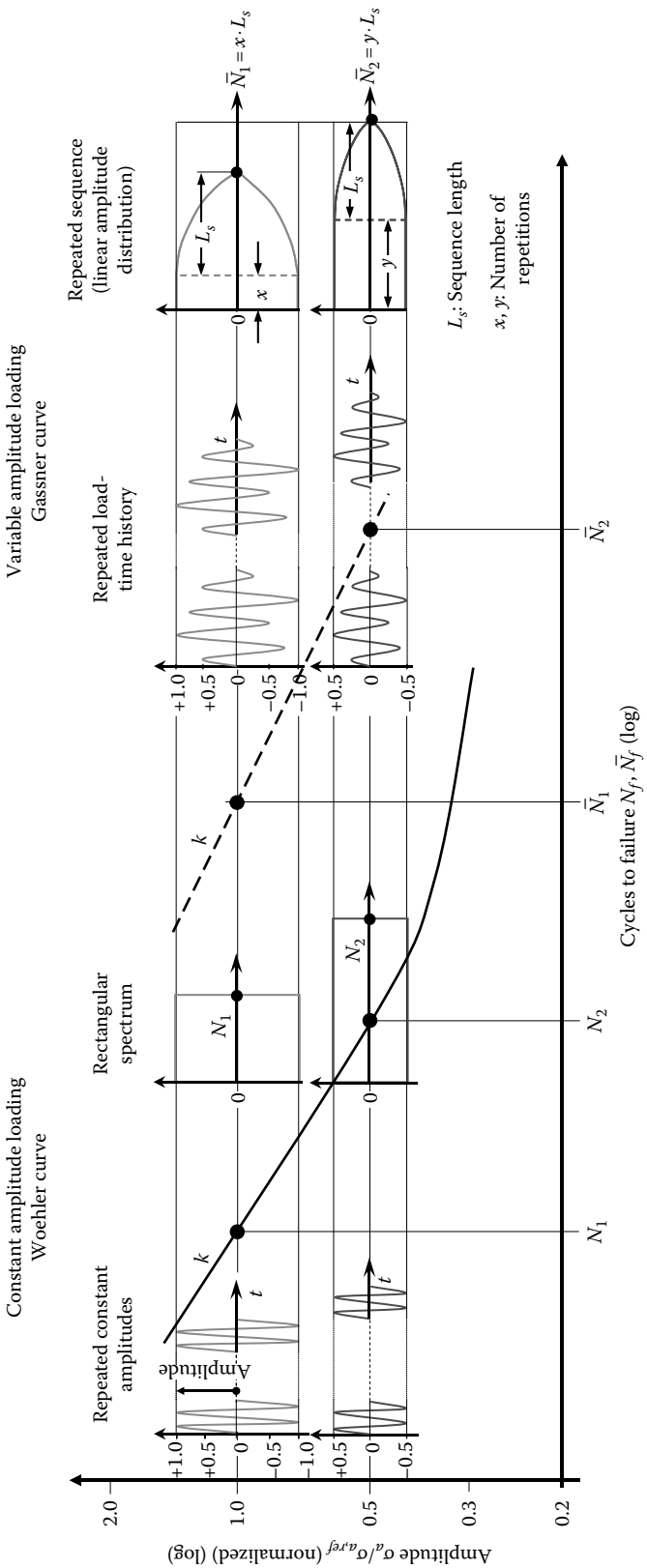


FIGURE 9.7 Presentation of constant and variable amplitude test results.

deformation a material cyclically hardens, softens, or is unaffected. Moreover, these curves are used for computing the notch root strains and stresses, for example, by implementing the Neuber rule or its modifications [20,21].

9.1.2.3 Significance of Tests with VAs

The significance of tests with VAs lies in the fact that, in contrast to tests with CA and identical peak values, they yield distinctly longer fatigue lives and exceed the Woehler curve in dependency of the spectrum shape. This is attributable to the fact that the smaller spectrum amplitudes in the Gassner tests damage less than the unchanged magnitudes of amplitudes continuously occurring in the Woehler tests. How much greater the fatigue life in the Gassner test is compared to that in the Woehler test depends on the shape of the sequences used. In the case of a straight-line distribution, the portion of smaller amplitudes is greater than those in the normal distribution (Figure 9.8); correspondingly, the associated Gassner curve shows a longer fatigue life. The advantage of knowing the spectrum's shape and its influence on the fatigue life for a given loading can be exploited for a lightweight design of component.

If, for example, a fatigue life of 10^8 cycles (after omitting the non-damaging load cycles) is required for a vehicle's steering rod and the peak load of $\bar{F}_a = 100$ kN measured in service is known, a diameter of $d = 22$ mm can be computed from the Woehler curve with an endurable stress amplitude of $\sigma_a = 260$ MPa. However, since a steering rod in service is subjected to a loading with VAs (for this particular case, measurements support a rectilinear distribution), knowledge of the corresponding Gassner curve for a fatigue life of 10^8 cycles with a stress amplitude of $\bar{\sigma}_a = 500$ MPa as a spectrum peak value, a diameter of 16 mm is computed. By allowing the Woehler curve for 10^8 cycles to be exceeded by a factor of 1.9, a reduction of cross section from 22 to 16 mm is achieved [22]. The consideration of VAs in engineering practice is the merit of E. Gassner (1939) and the most substantial improvement in lightweight component design, since A. Woehler (about 1870).

9.1.2.4 Size Effect and Transferability

Size effect is an essential factor in transferring fatigue data from specimens to components. The size effect can be divided into technological, stress gradient–related, statistical, and surface state–dependent contributions; however, a complete uncoupling of all size effect mechanisms is not

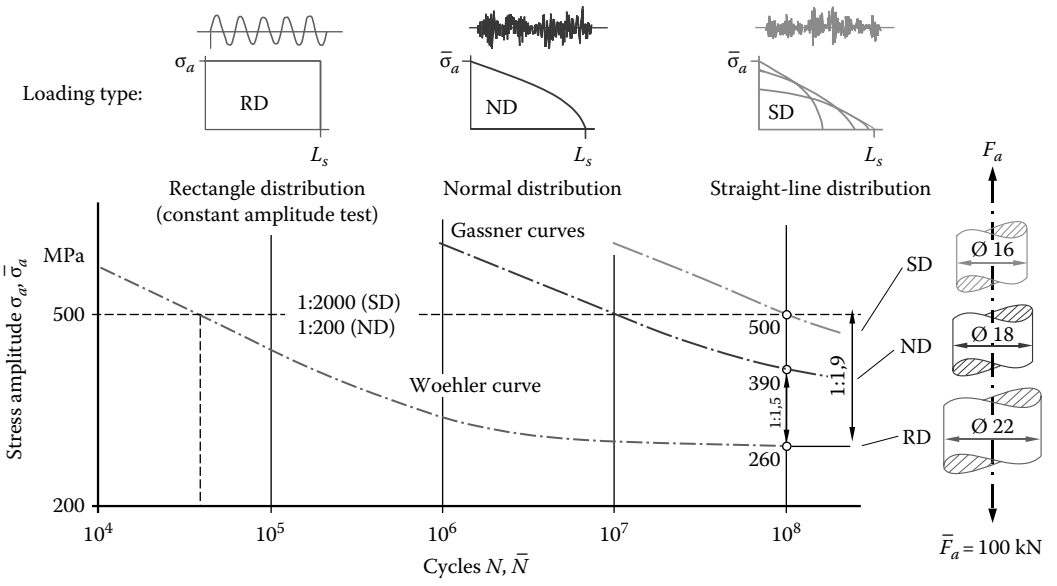


FIGURE 9.8 Influence of the spectrum on the fatigue life and component's dimension.

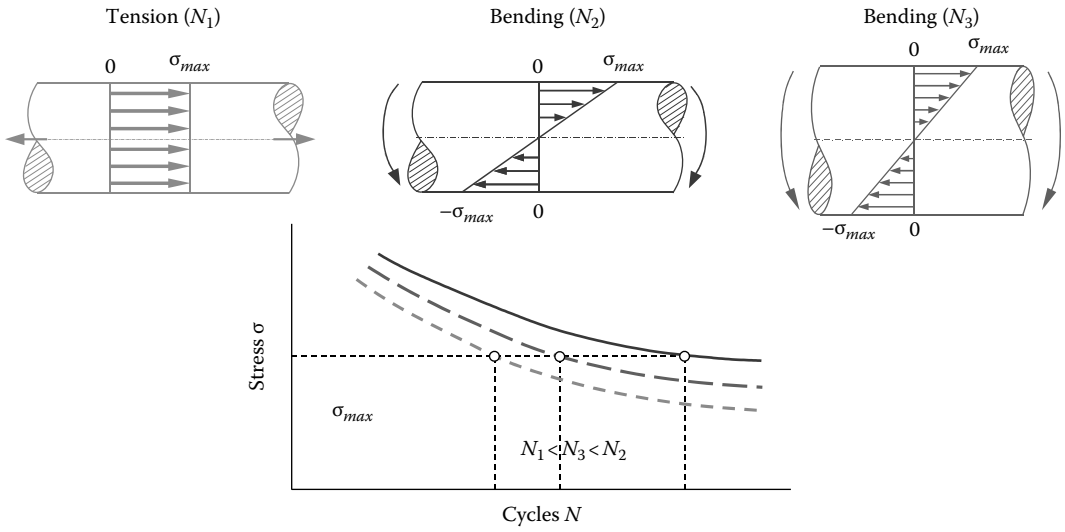


FIGURE 9.9 Influence of the maximum stressed material volume on the fatigue strength.

always possible [23–25]. A possible engineering consideration is given by the concept of highly stressed material volume, which considers the interaction between the stress gradient–related and the statistical size effect [20,21,26].

The highly stressed material volume is determined by the stress gradient depending on the component's geometry and the type of loading. The higher this volume, the higher is the failure probability. The so-called technological and surface state–related size effects are considered by separate examination of the surface, surface layers, and microstructural condition. Since the crack initiation not only is limited to the surface but also occurs in the region near to the surface, it has proved useful to define those volumes in which the maximum local stresses at the surface as well as into the depth fall by 10% (so that at least 90% of the maximum local stresses prevail) as the highly stressed material volume $V_{90\%}$ and to consider them for the evaluation purposes. The larger the highly stressed volume is, the greater is the probability that a material defect initiates a crack. For this reason, the local endurable stresses from larger highly stressed material volumes become smaller (Figure 9.9). For the transferability, it is initially assumed that the material, surface, and surface layer conditions, including the possible existence of residual stress conditions of, for example, specimens and components, remain the same. That is, the technological and surface state–related size effects are considered.

9.1.3 Component Design/Assessment

9.1.3.1 Initial Design/Assessment Data

In assessing and designing components, their functionality and their service behavior must be reliable.

For this assessment, many boundary conditions such as initial data must be known, for example,

- Loads (forces, moments, pressures, etc.) or local strains/stresses (including media and temperature), their magnitude and cumulative frequency (load spectrum), and their probability of occurrence
- Endurable stresses in the form of a fatigue life curve
- Design fatigue life and failure probability, etc.

Of special significance are the user-specific stress spectra. These are most appropriately determined by service measurements. The strength evaluation takes many forms and ranges from experimental proof tests on the component to pure computational verification (e.g., [27], Figure 9.10).

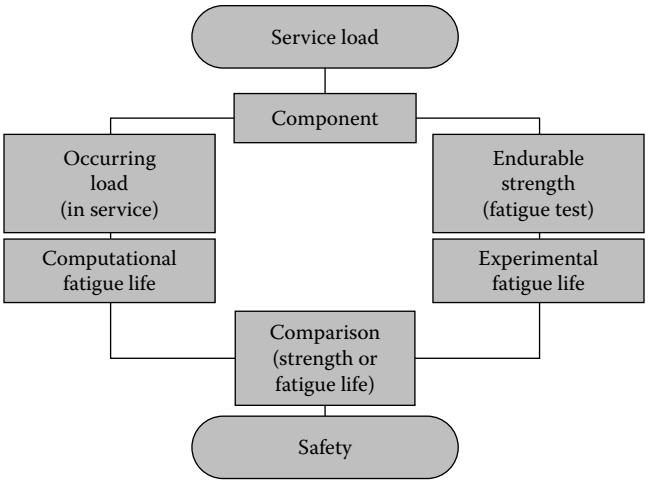


FIGURE 9.10 Component design (verification test).

9.1.3.2 Design Concepts

The criteria for an appropriate component design are the reliability of the components of the construction or of the entire system (service reliability) and the safety (no failures before reaching the designed fatigue life for a required theoretical failure probability). The following engineering design criteria are applied (Figure 9.11):

- Static design, usually against the yield strength.
- Design against the (so-called) endurance limit, assuming infinite life design, when the spectrum lies below the Woehler curve following the knee point, but the spectrum's peak value occurs much more than 10^6 times; this is the case, for example, with connecting rods, crank shafts, and gears. Although the loading occurs in these examples with VAs, they are treated as CA loadings because of the high occurrence of the spectrum's peak values.

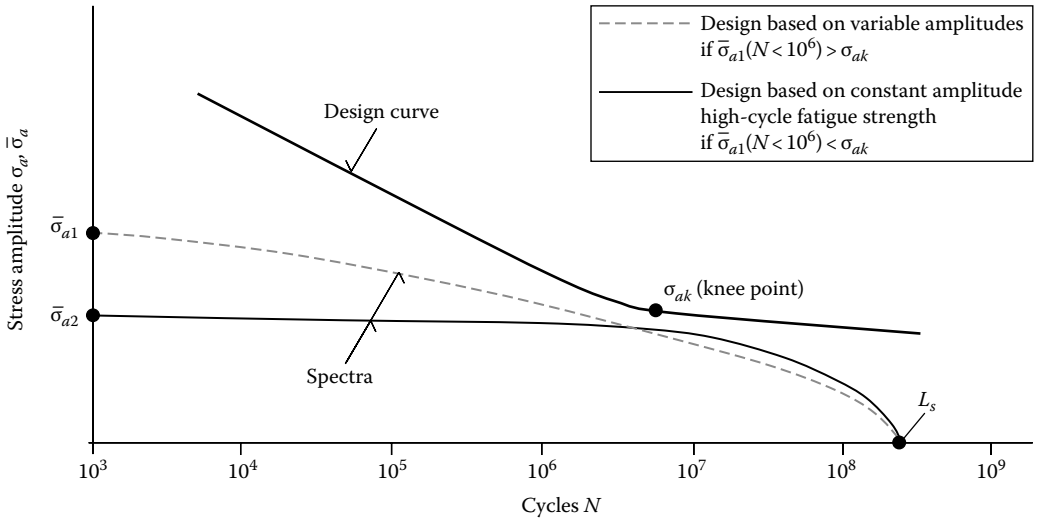


FIGURE 9.11 Criteria for component design based on constant and variable amplitude loading.

- Structurally durable design, when the spectrum's peak values exceed the Woehler curve following the knee point and also when a limited number of VAs are allowed. Permitting higher stresses than those at the Woehler curve's knee point leads to smaller component cross sections (lightweight design).

The following concepts are sometimes applied simultaneously in different technical branches:

Nominal stress concept: The nominal stress concept requires both the definition of a nominal stress and also the classification of a notch factor or, with welded joints, a fatigue class. According to the definition of both of these parameters, a design Woehler curve referring to the nominal stress system must be taken into account for which the same notch factor or fatigue class exists for the component that is to be evaluated. This concept, with which a nominal strain can also be used, reaches its limits of applicability for complex geometries for which neither a nominal stress nor a notch factor or fatigue class can be defined [28]. Moreover, since the present possibilities offered by numerical simulation (FEM) provide the local strains/stresses, this concept is increasingly superseded.

Structural stress concept: This was developed for welded joints [28]. The concept, which can also be based on strain, extrapolates the stress distribution outside of the weld's seam to the weld's toe according to stipulated criteria thus defining a hot-spot stress. This stress is then assigned a structural stress Woehler curve and then evaluated. The advantage of these concepts lies in the fact that for complex geometries, the problem of the definition of a nominal stress or that of a fatigue class is avoided. In contrast to the nominal stress concept, only few fatigue classes for welded joints are given. The disadvantage lies in the fact that critical regions, which are not externally accessible, for example, weld roots, cannot be assessed with this concept.

Local concept: The local concept can be used with both stresses (local stress approach) and strains (local strain approach). Here, the load-dependent notch stress or strain must be determined and for whose evaluation the local endurable stress or strain must be available. That is, the Woehler curves for the local system, taking into consideration the size effects, surface, and material state, must be available. The advantage of this concept, also used for welded joints, lies in the fact that difficult geometries can be assessed by calculating the load-dependent local stresses or strains.

Fracture mechanics approach: This concept [29] assumes that, in a component's critical region, crack-like defects or geometries exist, from which a crack can be initiated or propagated. Crack initiation can be treated by the definition of stress intensity and the crack propagation by additional linear-elastic or elastoplastic constitutive equations. If the crack initiation phase up to a defined length or depth is not taken into consideration, this part of fatigue life is not considered by this design method.

9.1.3.3 Strength Hypotheses

In general, components in service are subjected to multiaxial loads with existent and changing principal stress directions. For a fatigue life evaluation, the local stress states have to be reduced to an equivalent stress for comparing them with fatigue properties derived under uniaxial loading. For the evaluation of an equivalent stress, different hypotheses are established. They were primarily developed for static load cases, but some of them have also proved to be suitable for cyclic loading:

- Ductile materials (e.g., fine-grained steel, quenched and tempered steel, etc.): Shear stress hypothesis according to Tresca or the distortion energy criterion according to von Mises
- Brittle materials (e.g., cast iron, aluminium casting): Criterion of maximum normal stress according to Galilei

Proportional loading situations can be assessed fairly well with these criteria, but special attention has to be paid to changing principal stress directions, since materials react differently to this special load situation and the numerical assessment is highly complex. This is the case for ductile and

semi-ductile materials. Ductile materials reveal a decrease of fatigue life under multiaxial loading with changing principal stress directions in contrast to constant directions. Brittle materials show an increase of fatigue life while semi-ductile materials present an indifferent behavior [30,31]. To account for these different features, for ductile as well as for semi-ductile materials (e.g., cast steel, wrought aluminium), plane-oriented hypotheses have been developed in recent years. While for ductile materials the criterion for the calculation of an equivalent stress [30,31] is the integration of shear stresses over all planes, for semi-ductile materials the criterion is the critical plane with the most unfavorable combination of shear and normal stresses. In the case of brittle materials, the equivalent stress is the maximum normal stress occurring in the critical plane and corresponding to the principal stress. Also, approaches on the basis of strains, energy density, and short cracks [32,33] have been proposed but require detailed information of the material properties and are most demanding in application [34,35].

9.1.3.4 Estimation of Fatigue Life

The question of fatigue life estimation arises for VA loading [6]. To computationally estimate the fatigue life according to the hypotheses of linear damage accumulation [5] according to Palmgren (1924) and Miner (1945) (Figure 9.12), knowledge of the spectrum (Rainflow matrix or level crossings in combination with range pairs [16]) and also of the Woehler curve is required.

Since, in general, the spectrum's amplitudes show different *R*-values as displayed in a Rainflow matrix, a transformation on to the *R*-value of the Woehler curve has to be carried out. For the original Palmgren–Miner computation, the Woehler curve following the knee point is horizontal. That is, it is not considered that amplitudes below the knee point cause damage, which is wrong. Therefore, the Woehler curve, depending on the type of modification of the Palmgren–Miner rule, is continued after the knee point with a changed slope, for example, with $k' = k$ (Palmgren–Miner elementary) or with $k' = 2k–m$ (a modification according to Haibach; $m = 1$ for wrought materials, $m = 2$ for cast, welded, or sintered materials). This is carried out in order to allow for damage caused by amplitudes below the knee point. The most used modifications are aforementioned. Following the summation of the damage increments n_i/N_i, the fatigue life $\bar{N}_{calc}$ can be calculated (estimated) using the theoretical damage sum $D_{th} = 1.0$. However knowledge of the real damage sum D_{real} is required for a correct estimate which, in most known cases, is <1.0. The estimate can only be determined from tests (Woehler and Gassner curves) as a ratio of the experimentally determined fatigue life and that calculated with $D_{th} = 1.0$ ($D_{real} = \bar{N}_{exp}/\bar{N}_{calc}$ ($D_{th} = 1.0$)), and is linked to the computing procedure. In the event that other spectrum shapes exist in service as those used in the test, the position of the corresponding

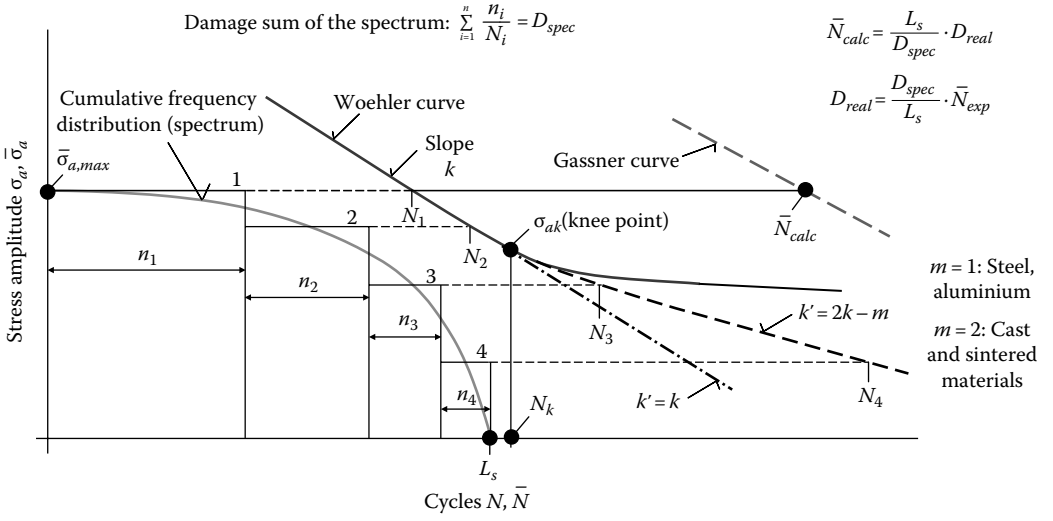


FIGURE 9.12 Modification of the Woehler curve and calculation of fatigue life (schematically).

Gassner curve can, in this case, be estimated by a relative damage calculation; however, it may not be always correctly calculated. The real damage sums can be scattered over three logarithmic decades. Approximately 90% of the experimentally derived damage sums lie below the theoretical value $D_{th} = 1.0$; that is, calculations with $D_{th} = 1.0$ give, in most cases, an overestimate of the fatigue life [36]. Therefore, the allowable damage sums $D_{all} = 0.5$ for welded joints and $D_{all} = 0.3$ for non-welded components (FKM guidelines [27]) are assumed for an initial assessment for estimating the fatigue life. For load-time histories with high variations of mean values, damage sums of 0.2 and 0.1, respectively, should be used instead of 0.5 and 0.3 [13,14]. Owing to these uncertainties, experimental verification of structural durability, where possible, is always recommended, particularly for safety critical components. In cases that an experimental verification is not possible, inspections are necessary.

9.1.4 Endurable Loading

9.1.4.1 Fracture Behavior Subject to Impact Loading

For a structurally durable design, especially safety critical components having developed unanticipated cracks must withstand loading due to misuse, for example, an impact, and must never reveal a brittle failure. Such loadings can occur while driving over obstacles, by potholes, or from a collision with a curb stone.

Figure 9.13 contrasts two commercial vehicle's wheel hubs made from forged and cast materials of the same weight. In both hubs, an artificial crack of 0.5 mm deep was inserted at the highly

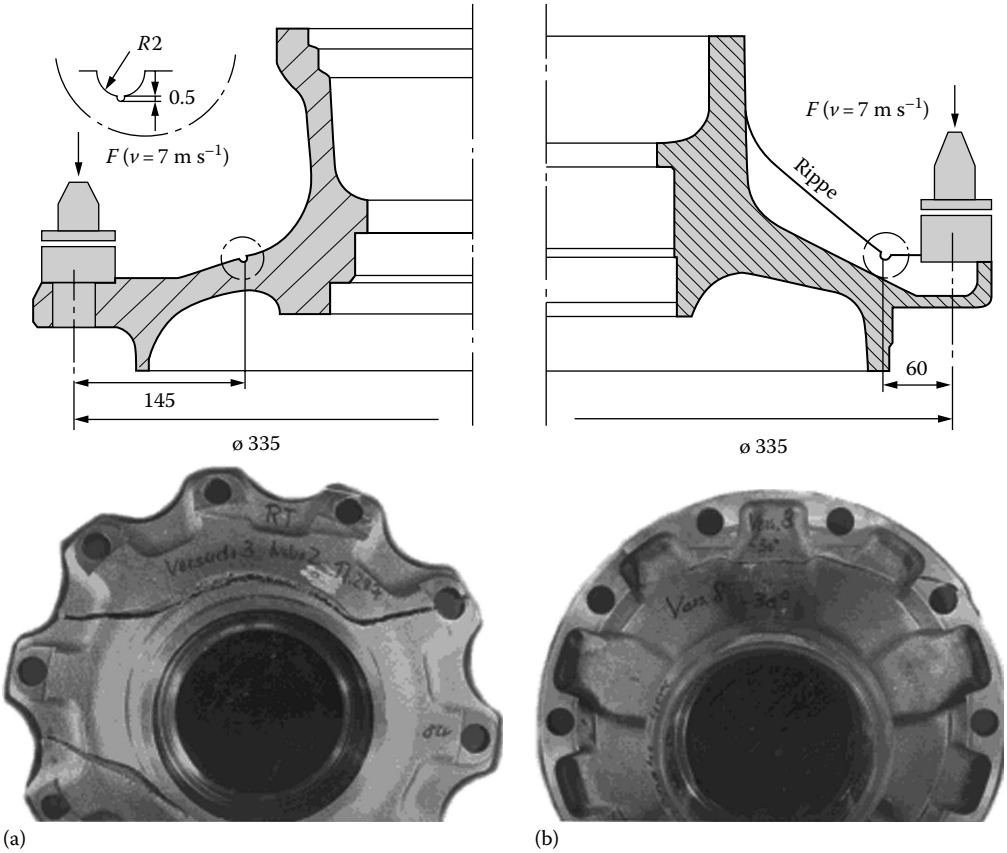
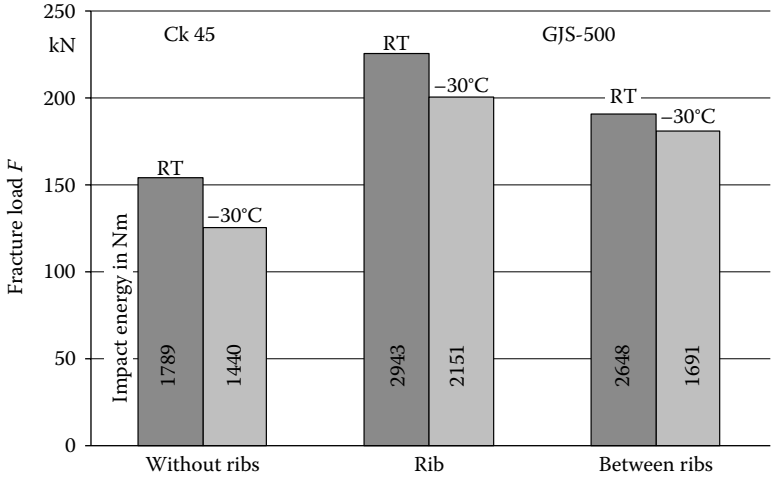


FIGURE 9.13 Influence of design on impact behavior of truck hubs with artificial cracks. (a) Forged hub of Ck 45, (b) cast hub of GJS-500.

(*continued*)



Material	$R_{p0.2}$ [MPa]	R_m [MPa]	A_e [%]	Z [%]	$A_v(RT)$ [J]
Ck 45 coarse-grain annealed	359	707	15	33	19
GJS-500	373	577	7	7	5

(c)

FIGURE 9.13 (continued) Influence of design on impact behavior of truck hubs with artificial cracks. (c) Fracture load.

stressed area. Using the conventional material data, the forged steel appears to be quite superior to the cast material, although, on the basis of the cast form (ribbing in the critical region), a significantly higher impact load is required to cause rupture in contrast to the forged hub. In this example, the magnitude of the rupture force is strongly influenced by the geometry. Thus, no correlation exists between conventional static and impact material data obtained with simple specimens.

9.1.4.2 Structural Yield Strength

The 0.2% yield strength of a material is the stress at which a remaining plastic deformation of $\varepsilon_{pl} = 0.2\%$ occurs and is determined from the stress–strain curve using an unnotched specimen ($K_t = 1.0$) in a tensile test or during a strain-controlled fatigue test. As soon as the linear elastic region of the stress–strain curve is exceeded, the material will tend to flow in an uncontrolled under manner axial loading of an unnotched specimen. The material flows according to its hardening properties in the absence of existing support effects (no stress gradients). This can only be prevented by strain controlling (strain Woehler curve), that is, the strain control simulates the flow restricting effect of strain–stress gradients in notches. On the other hand, in contrast to the 0.2% yield strength of an unnotched specimen, a higher local total strain is required to obtain a remaining plastic deformation of $\varepsilon_{pl} = 0.2\%$ by reaching the 0.2% component's structural yield strength [10] at a component or a notched specimen. This is caused by the stress gradients at the component's or specimen's notch. The ratio between the local total strain in the notch of the component and that at the unnotched specimen yielding a 0.2% remaining plastic deformation is defined as the support factor. The stress gradient in the notch, as long as the 0.2% component yield strength is not exceeded, locally impedes an uncontrolled yielding. That is, the gradients locally exert, despite the external

load control, a strain control. For this reason, there exists a greater linear region between external loading and local straining in components and notched specimens than in unnotched specimens. From the point of view of structural durability, the important component structural yield strength, which can be reached by a service stress as a consequence of a special event, must not impair the fatigue strength. According to material and component geometry, the allowable component yield strength may be lower than the 0.2% limit, for example, $\varepsilon_{pl} = 0.02\%–0.05\%$.

9.1.4.3 Choice of Relevant Design Data

Fatigue data are predominantly determined and made available using specimens. The informative value of these data depend to what extent the user, to whom the data have been made available, is in a position to evaluate the different factors of influence. Furthermore, the failure criterion is important. Fatigue data determined on unnotched specimens mostly provide a conservative approach, since the uniaxial loaded specimens possess no stress gradients. Higher strength values are determined under bending depending on the stress gradients and size of the specimens (Figure 9.9). It is, therefore, very difficult to make data for dimensioning purposes available, as displayed in Figure 9.3, since these always require a lot of knowledge (consideration of local properties like microstructure, strength, surface states, highly stressed material volume, surface treatment, etc.), which enables the reliable transferability of data from specimens to components. As long as no experience with respect to the transferability of fatigue data exist, a verification test on the actual component is always recommended. If this is not possible, a fatigue life curve should be derived using a specimen with comparable local properties taken from the component.

9.1.4.4 Safety Considerations—Endurable and Allowable (Permissible) Stresses

Starting from a fatigue life curve with a probability of survival of $P_s = 50\%$ (endurable stresses or strains), a design fatigue life curve (allowable stresses or strains) is derived for a component design. With regard to this, a theoretical probability of failure $P_f = 1 - P_s$ is initially defined according to the component's risk potential. For vitally important components, which must never fail, for example, brakes, $P_f \leq 10^{-6}$ (Gaussian log-normal distribution assumed). For components which are frequently inspected and which are operated in secure areas, for example, a pressure vessel in a bunker, $P_f = 10^{-3}$. Furthermore, the following must be known:

s_X: The standard deviation around the mean value $X(P_s = 50\%)$ of the endurable stress, strain, load, deformation, pressure with $s_X = (1/2.56)\log(1/T_X)$

s_M: The standard deviation of the mean value depending on the batch and production (quality control).

s_L: The standard deviation of the external loading; in cases where the maximum loading is assumed to be known, then $s_L = 0$.

By assuming a logarithmic normal distribution, the total standard deviation

$$s = \sqrt{s_X^2 + s_M^2 + s_L^2}$$

can be calculated from the individual standard deviations [37]. The allowable (permissible) stress, strain, etc. for a required theoretical failure probability is given by

$$X_{all} = \frac{X\left(P_s = 50\%\right)}{j_X}$$

$$j_X = 10^{-s \cdot u_0(P_f)}$$

$u_0(P_f)$ depends on P_f and is the normalized safety factor [3,4]. Table 9.1 shows the value of u_0 for several probabilities of failure P_f.

TABLE 9.1
Calculatory Probability of Failure P_f and Normalized Safety Factor u_0 for Gaussian Logarithmic Distribution

P_f	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
$-u_0$	1.28	2.33	3.09	3.72	4.27	4.75

A further good approach to obtain j_X is also given by [38]:

$$j_X = 10^{\left(2.36\sqrt{|\log P_f|}-1\right)s}$$

As the scatter is dependent on the stress level and the corresponding number of cycles, different magnitudes of safety factors will result along the fatigue life curve ($P_s = 50\%$) from which the design fatigue life curve is derived. The emphasis on theoretical failure probability is well founded, since the real distribution outside the regions $P_s \approx 5\%–95\%$ is unknown. Nevertheless, the implementation of the theoretical failure probability is legitimate when its specification is supported by verifiable experience with respect to the safe service of a component. To derive the safety factors, other distributions, for example, Weibull, can also be used. For this purpose, the corresponding service experience must also exist.

9.1.4.5 Experimental Proof Tests

Basically, components must be designed such that the required fatigue life is reached without failure. This requirement can be fulfilled according to the component's relevance using different quality and economic standards. These standards depend on the classification into primary and secondary components according to considerations of safety and service reliability; in this connection, primary components are classified as safety and functional components. For safety components, for example, steering knuckles, wheels, brakes, failure must be excluded, since this represents a danger to the life of the user and the environment. High safety factors underlie the design of such components. The avoidance of failure is indeed equally the aim for functional components, since the functionality of system will be then prevented (e.g., connecting rod and crank shaft). However, a failure should not cause a danger to the user and environment. On the other hand, damage to secondary components, for example, exhaust manifolds, have, in many cases, no direct effect on safety or functional reliability. As a result, higher quality requirements are set for primary components due also to the economic consequences of failure, for example, given by recall actions. For example, the dimensioning of vitally important safety components must cover all service loads according to the present state of technology and knowledge. For this, not only is the accurate acquisition of the service loads required, but also knowledge of critical regions for failure of the components and the relation between occurring and allowable equivalent stresses depending on manufacturing process, material, and component geometry is required. Accordingly, the tolerable quality of the component can be stipulated under consideration of the economical-efficiency. Quality requirements, which are coupled to the safety of a component, can only be verified by achieving the corresponding experimental proof tests under simulated service conditions in which critical regions and the type of failure must be recognized. A further advantage of fatigue tests lies in the fact that they constantly verify the stability of the production quality and are subjected to the use of statistical methods, for example, as is the case of the production of mass products like connecting rods and crank shafts. The proof of fatigue strength must be supported by the corresponding quality specifications in the construction, the material procurement, and the manufacturing. The measurement of service loads for the verification and, if needed, timely corrections of computations should, as far as possible, be provided. To shorten the test's duration, an omission (leaving out non-damaging low cyclic

loads) may be carried out under consideration of the appertaining amplitude's mean values and the damage equivalent compensation of omitted loads [39]. With multiaxial tests, the correlation between the individual time signals must be taken into account [40,41].

9.1.4.6 Notations

A_e	Elongation at fracture
A_v	Notch impact energy
D	Damage sum
d	Diameter
F	Load
L_s	Sequence length
j	Safety factor
k, k'	Slopes of fatigue life curves
K_t	Notch factor
m	Wild card for calculating k'
M	Mean stress sensitivity
n	Cycle numbers of spectrum segments
N	Cycles
P_f	Probability of failure
P_s	Probability of survival
R	Ratio (min/max of load, stress, or strain)
$R2$	Radius of 2 mm
R_m	Tensile strength
$R_{p0.2}$	Yield strength
s	Standard deviation
T	Temperature
T_σ, T_N	Scatter in stress direction or cycles direction
U_0	Normalized safety factor
V	Loading velocity
$V_{90\%}$	Highly stressed material volume
X	Wild card for stress, strain, load, deformation, pressure
Z	Area reduction at fracture
Δ	Range
Δs	Deflection
α	Angle
ε	Strain
σ	Stress

Indices

1, 2, 3, 4, or i	Numbers for differentiation
a	Amplitude
all	Allowable
calc	Calculated
exp	Experimental
f	Failure
k	Referring to the knee point
L	Referring to the loading
m	Mean
M	Referring to the mean value
max	Maximal
min	Minimal

n	Nominal
pl	Plastic
real	Real
spec	Spectrum
th	Theoretical
x	Wild card for N or σ
X	Wild card for stress, strain, load, deformation, pressure

9.2 LOAD DATA ACQUISITION

Measurements form the basis of the operating design of almost all mechanical engineering safety components. In particular, knowledge of the magnitudes and frequencies of the applied loads and of the stresses and strains during the operational life required for the customer's application are essential as input parameters for evaluating structural durability. In order to determine these data for vehicles and machines, applied loads and local stresses and strains are measured during operating tests over relatively short time periods of 1–4 weeks. These data can then be extrapolated to the required fatigue life in order to

- Carry out an appraisal of the fatigue life using the applicable regulations
- Derive a testing program
- Make the input parameters available for a numerical simulation

According to each problem definition, strain gauges are located at different places. Loads are measured at the least disturbed regions. Critical strains are determined in regions where the highest stresses are expected. In Section 9.2.1, details of the load acquisition are dealt with and in Section 9.2.2, the application of strain gauges at critical regions is addressed.

9.2.1 Acquisition of Load Data

Loads form the input parameters, which ultimately cause the damaging stresses and strains in components. By considering road vehicles, wheel loads have over the last 30 years been determined as the global input parameters. Initially, chassis components were instrumented and calibrated. Today, measuring wheels are available for road vehicles, with which the three components of both force and moment at the wheel–road contact point can be determined. For rail vehicles, there is, at present, no satisfactory solution. A measuring wheel developed for this purpose for local track transport [42] is depicted in Figure 9.14. One can see that the wheel design incorporates pockets for locating ~100 strain gauges. Here, the difficulty lies in the positioning of the strain gauges such that the three components of both force and moment can be measured with the least possible cross talk. Simultaneously, the structure is to be sufficiently compliant such that the strains lie within the measurable range, but the fatigue strength of the structure is noncritical for the intended application.

To determine internal loads in systems, load sensors are integrated into the line of force transmission. In particular, of great interest are the internal loads in over-determined systems, in which elastomeric bearings are integrated—for example, double wishbone axles in automobile chassis. Here, system components are then, as a rule, instrumented with strain gauges and calibrated for force or moment (Figure 9.15). However, this has the disadvantage that they can only be applied for that specific system. The complexity of the problem definition increases with the number of degrees of freedom to be measured. Nevertheless, since the quality of the numerical multibody simulation steadily increases, the data acquisition technology for measuring loads are, at present, becoming increasingly more important for structurally durable designs.

The great advantage of measuring load data in comparison to local stresses and strains is that the former are independent of component geometry. By measuring load data, a system is described with which the results can be transferred to modified or similar systems. Thus, for the

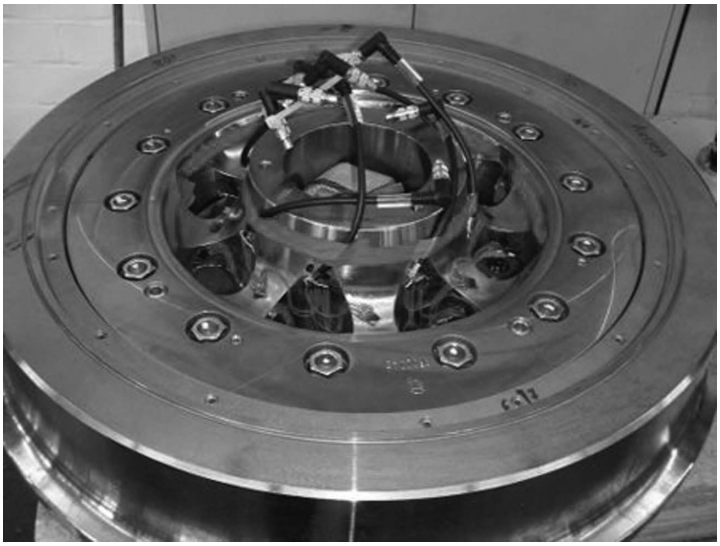


FIGURE 9.14 Tram wheel load sensor with machined pockets and strain gauges.



FIGURE 9.15 Instrumented and calibrated spherical hinge bearing as a load sensor for two load directions.

vehicle- and machine-manufacturing industries, load measurements form the basis of the continuous development of loading assumptions. This advantage can be used to increase the reliability of loading assumptions for designs, in particular the design of safety components. Moreover, by using long-term measuring concepts—the so-called "usage monitoring"—the changing usage profiles and customer requirements can be initially determined or finally verified.

9.2.2 Data Acquisition of Stresses

If strain gauges are implemented to evaluate fatigue strength, then the choice of sensors, their location, and orientation are of great importance. To ensure that the usual stress–strain relations and component fatigue strength are not impaired by holes, notches, keyways, and cut-outs, strain gauges can in general only be manually attached to accessible locations on the accessible surfaces of the component. If the governing stresses and strains are to be directly measured at a critical location, then the location of the measurement must first be chosen to be as near as possible to the assumed initiation point of the fatigue fracture. Secondly, the direction of the measurement must be chosen to coincide with the maximum principal strains and, if required, with the transverse direction. Since welded seams exhibit different characteristic types of fractures and failure locations, the correspondingly different possible strain gauge configurations are to be distinguished.

In joints that fracture from an undercut notch at a transitional cross section (see Figure 9.16), the location of the fracture initiation is, as a rule, accessible such that a strain gauge can be suitably attached near to the fracture initiation point.

In this case, the strain gauges are applied to the parent material at a definite distance from the seam. The reason for this is that the strains, which are integrated over the gauge length of the strain gauge and since, in the region of the run-out or in the notch of the seam, the cross section is not defined; a small shift of the strain gauge could produce a large change in the measured strain. The choice of the gauge length of strain gauge and its distance from the run-out of the welded seam depends on the plate thickness concerned. In Figure 9.17 and Table 9.2, recommendations are given for the choice and location of the strain gauges. Similar procedural regulations are given in the technical literature, for example, applications according to the "hot spot" structural stress concept of the IIW guidelines [43].

In joints where seams are subjected to longitudinal stresses and where the fracture runs transverse to the seam, one may assume in many cases that the overall fractured cross section was subjected to a uniform longitudinal strain. For this type of joint, the strain gauges that are attached to the parent material next to the run-out of the seam in the direction of the seam give, as a rule, the strain at the fracture initiation point (Figure 9.16c and d). In particular cases, it is advisable to attach a second strain gauge to the opposite side of the sheet or seam. By assuming a linear strain distribution, it is then possible to interpolate the strain at the fracture initiation point from the strains at both measuring locations.

In contrast to this, joints, whose fracture initiation points are located at notch roots, are in general inaccessible to a direct stress–strain measurement. In these cases, knowledge of the notch root geometry is required in order to construct a numerical model of the welded seam. This numerical model can be verified using data from strain gauges attached near to the welded seam subjected to

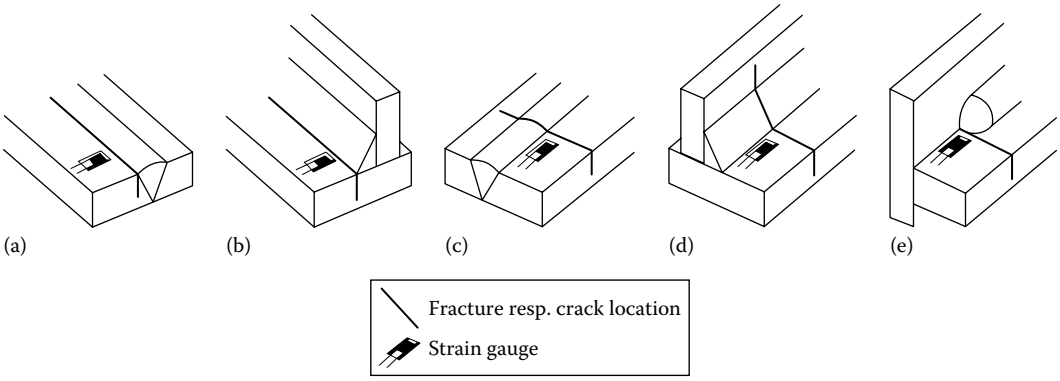


FIGURE 9.16 Possible configurations of strain gauges at welded seams where the fatigue fracture is initiated on the seam's upper side, that is accessible to measurement.

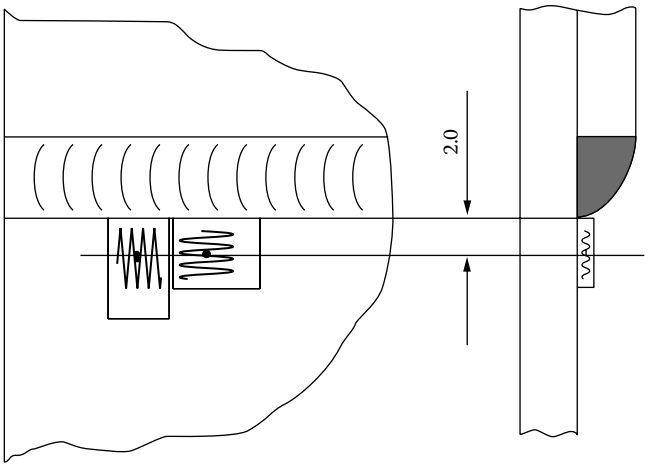


FIGURE 9.17 Recommended applications of strain gauges at run-out of welded seam.

TABLE 9.2
Recommendations for Attaching Strain Gauges at Run-Out of Welded Seam

Sheet/Plate Thickness (mm)	Strain Gauge Length (mm)	Strain Gauge Distance-Center Line Seam Run-Out x (mm)
1–5	1.5	1.25
5–25	3.0	2.0

specified loads, which occur under operating conditions. The evaluation of the notch root is then carried out using the numerical simulations and local concepts.

9.3 DERIVATION AND GENERATION OF LOADS FOR EVALUATION AND TESTING

Realistic assumptions about loads are essential for a sound analysis of structural durability. A simple application of a CA loading instead of the loading during service, which is characterized by VAs can lead to misjudgements when analyzing complex components with several possible failure sites.

Figure 9.18 shows an example in which failures occur at different regions depending on the chosen amplitude of the CA loading. A fatigue analysis is only possible by considering the service load spectrum.

The scatter band of service loads is typically larger than that of the component's strength (see Figure 9.19). On the other hand, relatively small changes in the loading already lead to comparatively large changes in the fatigue life. For example, an increase in the load's amplitude of 15% leads to a shortening of the fatigue life by almost a half (assuming a CA loading and an S–N curve's gradient of $k = 5$). The outlay on more appropriate load assumptions reduces uncertainties in design and thereby leads to lighter or more cost-effective designs.

9.3.1 Fundamentals and Influences on Fatigue Life

In deriving the load-time series or load sequences, various aspects should be taken into account. Some of these aspects are discussed in the following. For an overview on load data analysis from the perspective of structural durability, see [44–51].

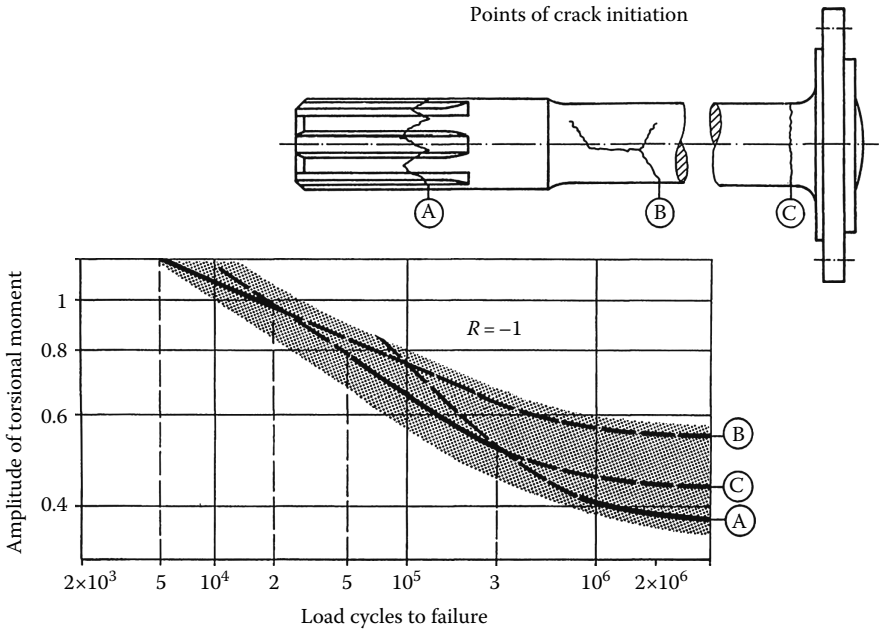


FIGURE 9.18 S–N curves of a drive shaft under CA torque loading: locations of crack initiation depend on load amplitude.

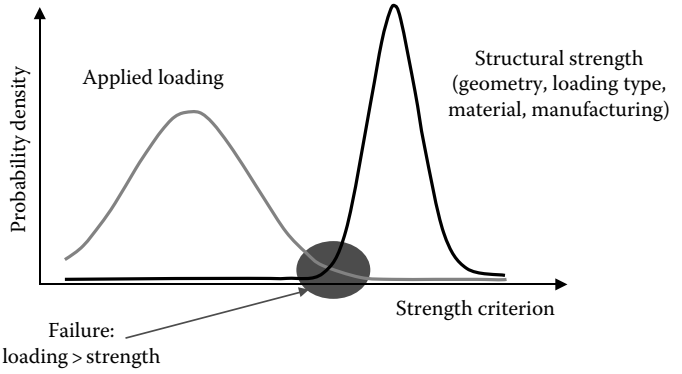


FIGURE 9.19 Applied loading and structural strength: 1D perspective.

The derivation and generation of loads for numerical and experimental analyses of fatigue life are, as a rule, based on evaluating service loads (e.g., using measurements of loads during service), and on a service profile, which is given by the anticipated use of the structure under consideration (see Figure 9.20).

Counting procedures [52] form the basis for the specification of service loads, that is, the determination of parameters (amplitude, mean value) and the number of occurrence of cycles. In the case of multiaxial loading or dynamically responding systems, other characteristics of the load-time function are important (correlation, frequency content, etc.). In each case, the operating loads are to be determined and specified separately for all loading conditions occurring during the anticipated use of the construction.

The service profile includes the frequency, distribution, and sequence of the individual loading conditions for the construction's service life. On establishing the service profile, decisions must be made that go beyond the purely technical considerations, since they can affect questions of product liability, costs, competitiveness, company image, etc.

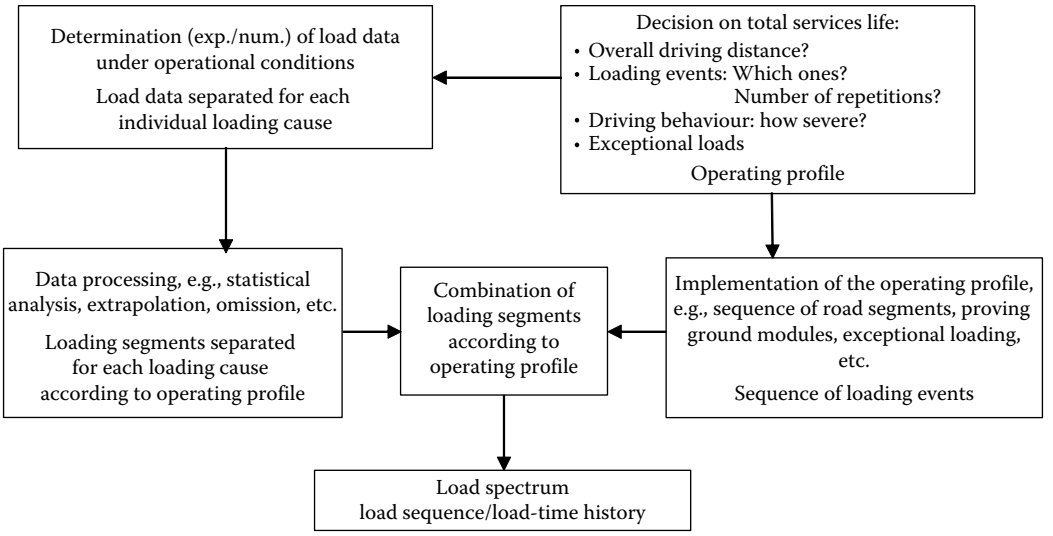


FIGURE 9.20 General approach for load assumptions: determination of load spectra or load sequences.

Both the compendium of loads, determined for all the individual loading conditions corresponding to the service profile, and the allocation of a probability of occurrence ultimately lead to a design spectrum which can directly lay the foundation of a numerical analysis.

Usually, the number of cycles in the design spectrum is so large that further modifications are necessary with regard to a reasonable test period for deriving test spectra or testing load sequences (drive files), as the case may be. Methods for generating shortened load sequences or load-time functions, as the case may be, will be elucidated.

9.3.1.1 Mission Analysis and Usage Monitoring

Typically, the total loading of a structure is generated by more than a single load or one operating condition. Examples are the different phases, maneuvers, payloads, and environmental conditions of a vehicle's usage (cornering, braking/accelerating, rough road conditions). The compilation of the overall spectrum requires one to separately record the individual operating conditions and to extrapolate them to the total service life. Figure 9.21 exemplifies the sub-spectra of the vertical and transversal forces, which were determined for a railway vehicle's wheel set (extrapolated to the total mileage with more than 10^9 cycles) [53].

The spectrum's region, which covers the most frequently occurring amplitudes, was obtained by superimposition of the sub-spectra multiplied by the number of repetitions during the total service life. Extreme values are to be considered separately.

For cars or motor cycles, the service profile of the customer's usage is not known completely. In this case, the basis for compiling loading spectra can be provided by questionnaires, long-term measurements on customer's vehicles as well as by carrying out supplementary measurements under known conditions [54].

9.3.1.2 Loading Sequence and Load Amplitude Variation

It is known that the sequence and the (missing) variation of load amplitudes influence the fatigue life. Figure 9.22 depicts an example of this for measured service loads. Experimental investigations on circumferentially notched specimens subject to synthetically generated load sequences (Figure 9.23) show differences in fatigue fracture life by a factor of $f_N \approx 0.5$–2, depending on the order of the load cycles. As a consequence, attention should be given to a realistic combination of the individual load amplitudes on generating load-time sequences.

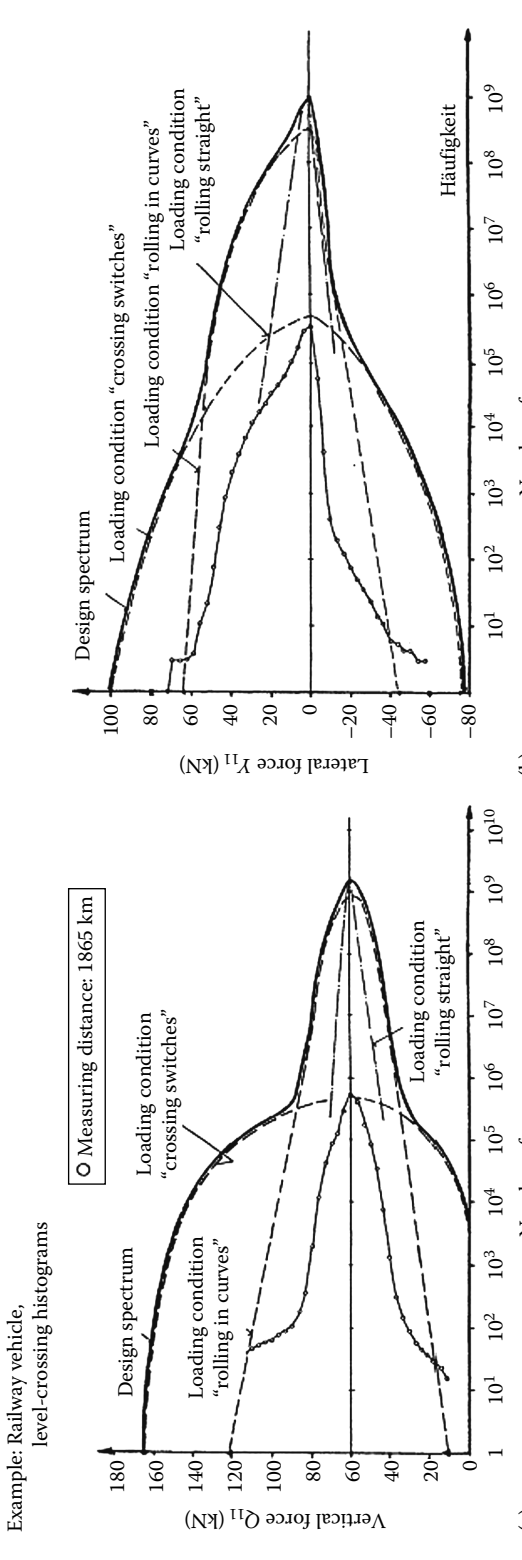


FIGURE 9.21 Composing a design spectrum based on individual loading conditions: (a) vertical force Q_{11} and (b) lateral force Y_{11}.

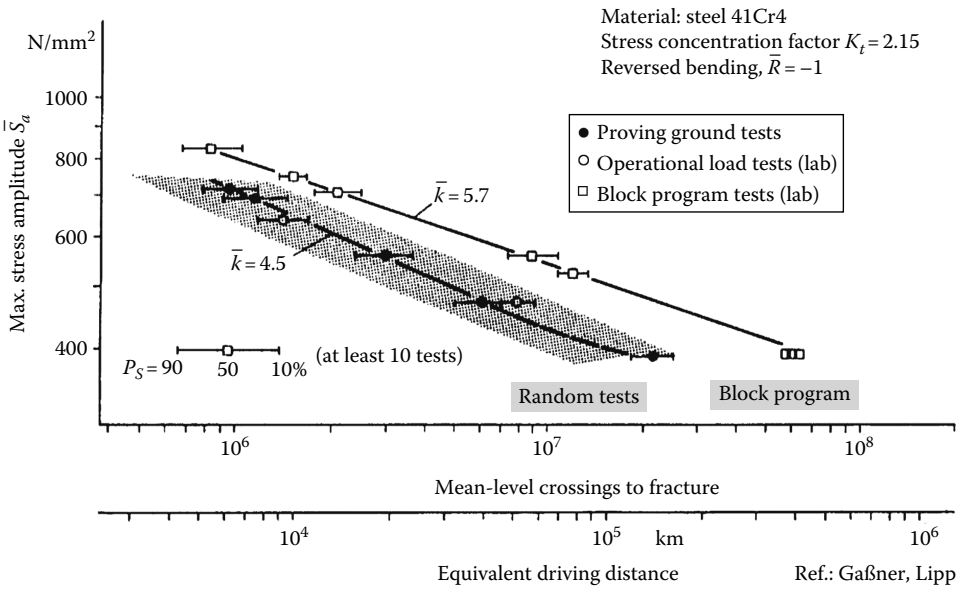


FIGURE 9.22 Fatigue life curves from proving ground and laboratory tests (constant calculated damage sum for same maximum stress amplitude).

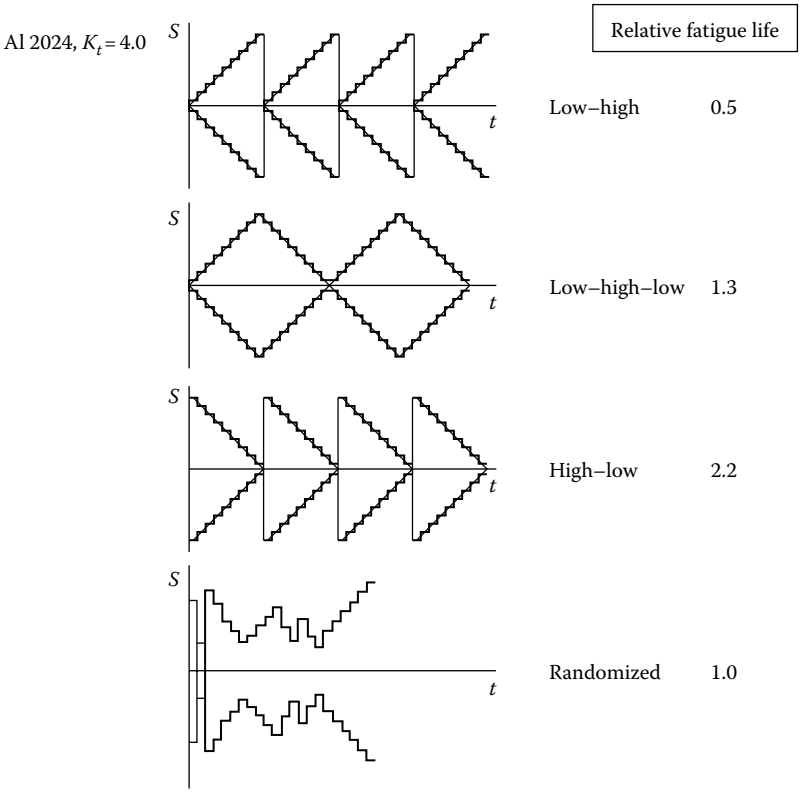


FIGURE 9.23 Influence of sequence and combination of cycles on the fatigue life of notched specimens.

9.3.1.3 Influence of Overloads on Fatigue

On specifying extreme loads, overloads that only rarely occur or special events for a testing sequence, different phenomena in material behavior are pertinent. The effects of extreme loads or overloads on the fatigue life depend on, among other things, the type of loading, magnitude and frequency of occurrence, component's geometry as well as the material. Overloads can both extend or shorten the fatigue life (or act in a neutral manner).

Shell-type structures in the aircraft industry mainly demonstrate fatigue crack growth under plane stress conditions. Here, the fatigue life–extending effect of (tensile) overloads is known to be related to crack growth retardation due to crack closure. This influence is less pronounced for compact and massive components (where plane strain conditions as well as a smaller plastic zone are more likely at the crack tip).

Larger pre-strains that exceed the elastic limit in material specimens lead to a decrease in fatigue strength for a high numbers of cycles applied. However, overloads frequently have less or no effect on the structural strength of metallic materials provided that the overloads produce no macroscopically observable cracks or stable residual stresses that are not rearranged owing to the operating loads. For extremely flat S–N curves (e.g., material specimens, lightly notched components subject to torque loads), overloads can significantly decrease the fatigue life or the fatigue strength, as the case may be.

9.3.2 Extrapolation of Measurements

Estimating expected extreme load values is an important aspect of specifying the (sub-)load spectra. If the expected extreme value is not already contained in the measurements (estimatable, e.g., from physically limiting conditions), extrapolations using extreme value statistics provide useful assistance [50]. Here, it is assumed that the frequency distribution of the process, from which the extreme value originates, obeys an exponential distribution law and that no physical limit exists for the maximum load values. Within the scope of the analysis, the question should be answered: how the scatter of the measured extreme values is to be treated. That is, whether the scatter of the measured extreme values (a random sample) and that of the design spectrum (total population) resemble each other.

In [55], a procedure is described in which only those extreme values (maxima or minima, as the case may be) which lie, respectively, above and beneath a stipulated threshold provide the basis for the extrapolation. This is particularly important when the data base contains only a small number of measurements. That is, more than one extreme value must be employed using a single measurement. The extrapolation can also be carried out based on the Rainflow matrix [56,57]. Here, the distribution of the additional numbers of cycles, which arise from the extrapolation, is described by a scalable function, the so-called kernel. The application of this procedure requires some experience, in particular for the scaling of the kernel.

9.3.3 Shortening of Load-Time Sequences

It is frequently required to shorten a design or test spectrum in order to be able to carry out tests within an acceptable time period and budget. Figure 9.24 gives an overview of the methods which permit the shortening of uniaxial load signals.

Typically, shortening processes that operate with time windows are applied to time sequences, since these processes also allow handling of multiaxial loading. Here, the damage-relevant time segments are to be identified in the individual channels, which may be supplemented with their linear combinations. The damage-relevant time segments are identified for each channel. Such time segments, which do not show any significant damage content for all individual channels, are removed from all channels. The Rainflow counting [58] is generally employed for components

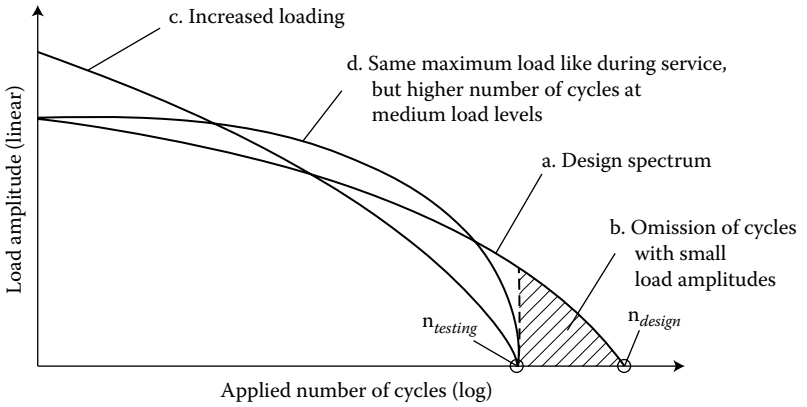


FIGURE 9.24 Overview on procedures to reduce the testing time.

whose fatigue life is independent of the loading frequency and can be characterized by the amplitude, the mean value and the number of cycles applied. With uniaxial loading, a shortening process can also be carried out based on a Rainflow matrix.

9.3.3.1 Omission

In the first instance, small cycles can, for this purpose, be removed or filtered (omitted) from the load sequence. Here, the criterion must be that neither the fatigue life (expressed as equivalent service life, driving distance, etc.) nor the failure site(s) are changed. The following have been found to be reference values for "permissible" filter limits for metallic components (e.g., chassis components with typical spectrum range of approximately $H_0 = 10^7$ cycles):

a. Approximately 15% of the maximum amplitude
b. Approximately 50% of the considered detail fatigue strength at the knee point of its S–N curve

For criterion (b), knowledge/assessment of the "fatigue strength at the knee point of the S-N curve" is required (the knee point of the S–N curve characterizes the strength, which does not, or only slightly, falls with higher number of cycles). On testing complex structures with numerous potential failure sites (e.g., a vehicle's front axle), this method is thus—without a previous numerical analysis—applicable only to a very limited extent.

It is advantageous to carry out the filtering using the Rainflow method in order to preserve the sequence of the remaining load cycles. Furthermore, the influence of the mean loads is taken into consideration (filtering out cycles of larger amplitudes with zero mean loads in comparison to cycles with nonzero mean loads). This requires knowledge of the mean load effects (e.g., the mean load sensitivity M). When the relationship between the load and the local stresses and strains is not known or not well-defined, both the tensile and compressive mean loads must be considered as detrimental. Bending load cases are an example of this.

For components which are multiaxially (nonproportionally) loaded, a shortening is typically carried out based on the load-time sequence. Simultaneously, additional channels, which are generated from linear combinations of the original channels, can also be taken into consideration. Using the procedure to simultaneously condense and optimize multiaxial time signals as an example [59], the modification of a five-channel signal for operating a test rig is outlined in Figures 9.25 and 9.26. Besides omission, adapting the load variations (velocities, accelerations) to the potential of the test rig can still be carried out. With respect to the latter, the component damage must be independent of the loading frequency.

FIGURE 9.25 Simultaneous procedure: example with five load-time histories.

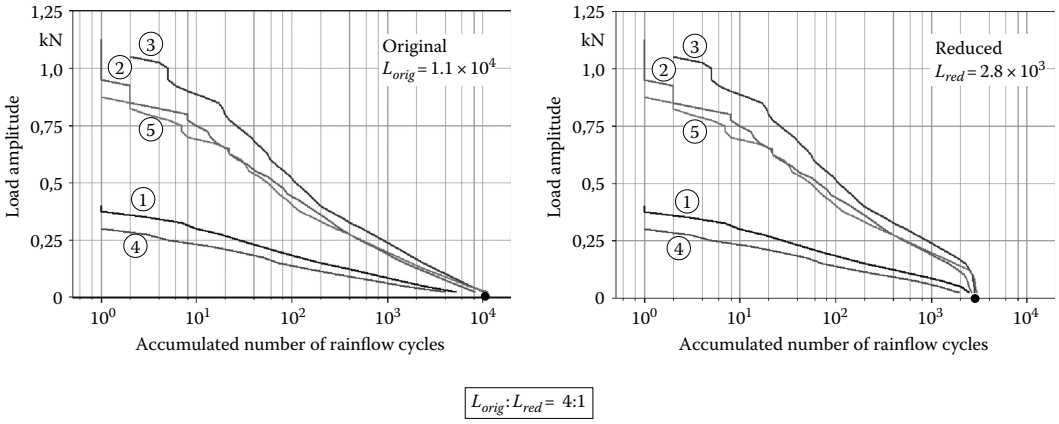


FIGURE 9.26 Simultaneous procedure: range-pair histograms of original and reduced signals.

9.3.3.2 Truncation

Limiting overloads or high loads (truncation) leads to a shortening of the fatigue life for shell structures (plane stress conditions), such as those prevailing in the aircraft industry. There is a danger of overestimating the fatigue life of aircraft parts, which are not subjected to these overloads in service. With larger sheet thicknesses, the crack retardation effect is not or, as the case may be, is less strongly pronounced. In massive structures, overloads, if anything, tend to shorten the fatigue life or act neutral manner.

Unrealistically high overloads or, as the case may be, unduly raising the total level of loads to shorten the testing time conceal the danger that this causes unrealistic rearrangements of the residual stress field or a loosening of bolts, which can change the test results compared to the actual service operations.

9.3.3.3 Elastomeric Components

Filtering load-time sequences for complex elastomeric components, such as hydro bushings, requires additional considerations. In contrast to the assessment of massive metallic components, the nonlinear force-deformation behavior as well as the frequency-dependent effects, such as for example the attenuation, should be considered in the load analysis. Based on investigations in [60], procedures [60–62] were introduced, in which the external loading was resolved into separate loading components (elastic force, damping force, and force at limit stops) by taking into account quasi-static load-displacement curves in order to improve the detection of different causes of damage. Here, the filtering may only be carried out in time periods where the deformations are not limited by contact.

9.3.4 Synthesis of Load Data

It is assumed that load-time segments (gained from measurements or numerical simulation) exist for both the relevant individual loading conditions (and channels) and the service profile. With this input, the design spectrum can be compiled. Examples of loading sub-spectra are depicted in Figure 9.27 for an axle component shown in Figure 9.28. The individual load-time segments contain the information concerning both the correlation of the channels and the frequency content during service loading.

9.3.4.1 Combination of Loading Segments

Typically, the load-time functions, which form the basis of the design spectrum, are quite long and therefore unwieldy for both the physical testing (drive-file iteration) as well as the numerical simulation, e.g., the simulation for multiple bodies. Thus, often a load-time sequence is compiled that consists of as few correlated load segments as possible, which are typical for the individual physical load cases. With this, the correlation (and frequency content) corresponding to the service conditions is guaranteed for the fatigue-critical loadings. An example of this is the

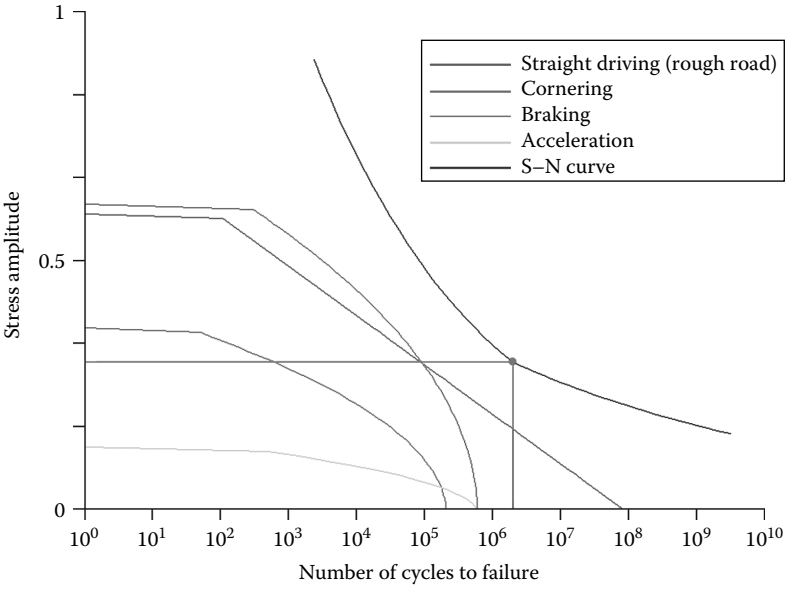


FIGURE 9.27 Local stress spectra for individual loading conditions.

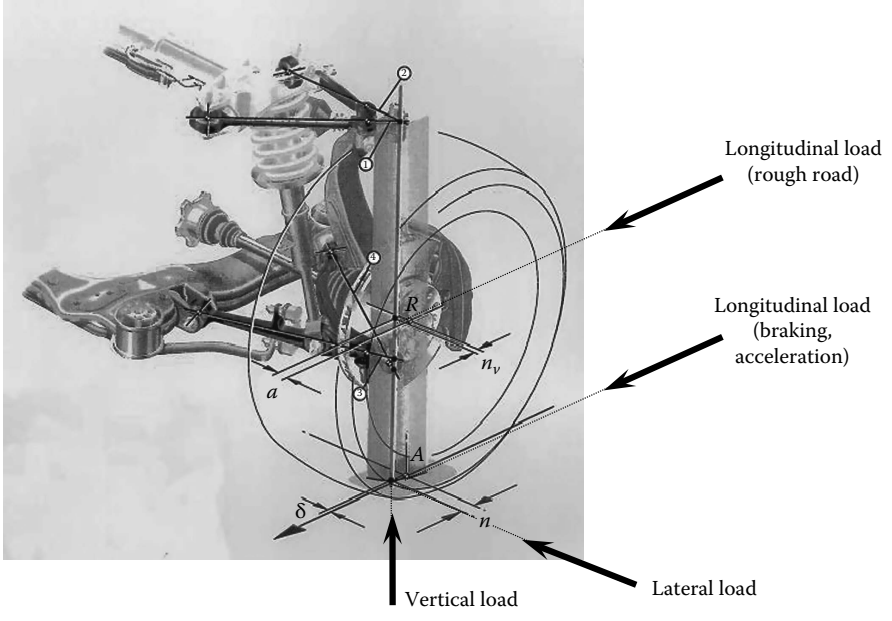


FIGURE 9.28 Multiaxial loading—example: loads acting at a wheel.

multiaxial loading program, which is based on synthetic segments for axle components outlined in Figure 9.29.

In the following, the general case of multiaxial loading is to be illustrated, for which a design spectrum exists in the form of a set of Rainflow matrices or range-pair and level-crossing histograms (for individual channels), and for which time histories for all channels must be derived. In the Simultaneous Procedure for producing multiaxial load-time sequences [59], the required load segments LS_i together with their number of repetitions (or frequency) $f(LS_i)$ and, if necessary, their

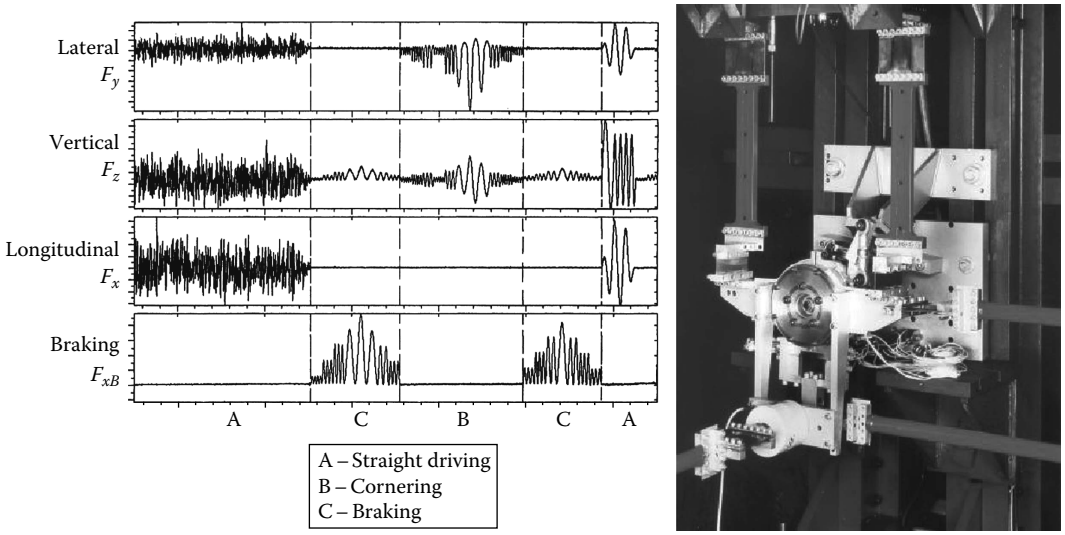


FIGURE 9.29 Multiaxial load program for axle components based on synthetic segments.

scaling factors $a(\mathrm{Ch}_j; \mathrm{LS}_i)$ are determined using an optimization procedure (Figures 9.30 and 9.31). A similar method is described in [63], which is based on an optimization of the Rainflow matrices without the option of scaling amplitudes. A prerequisite for the success of these approaches is that the basic load segments (events) contain all the relevant combinations (phase and amplitude relations) of the acting loads.

With uniaxial loadings, the synthesis of the load data can only be carried out using Rainflow matrices under specific boundary conditions. Here, the load sequence is derived from the resulting Rainflow matrix by means of a reconstruction procedure.

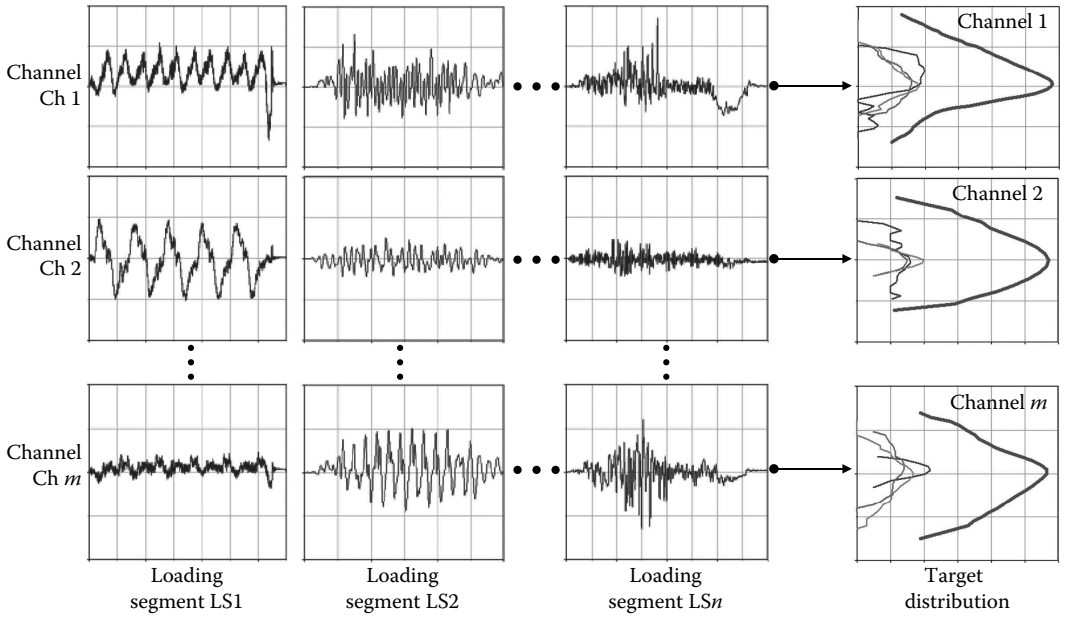


FIGURE 9.30 Generation of multiaxial load-time histories by correlated loading segments (determine frequencies $f(\mathrm{LS},i)$; scaling factors $a(\mathrm{Ch},j; \mathrm{LS},i)$).

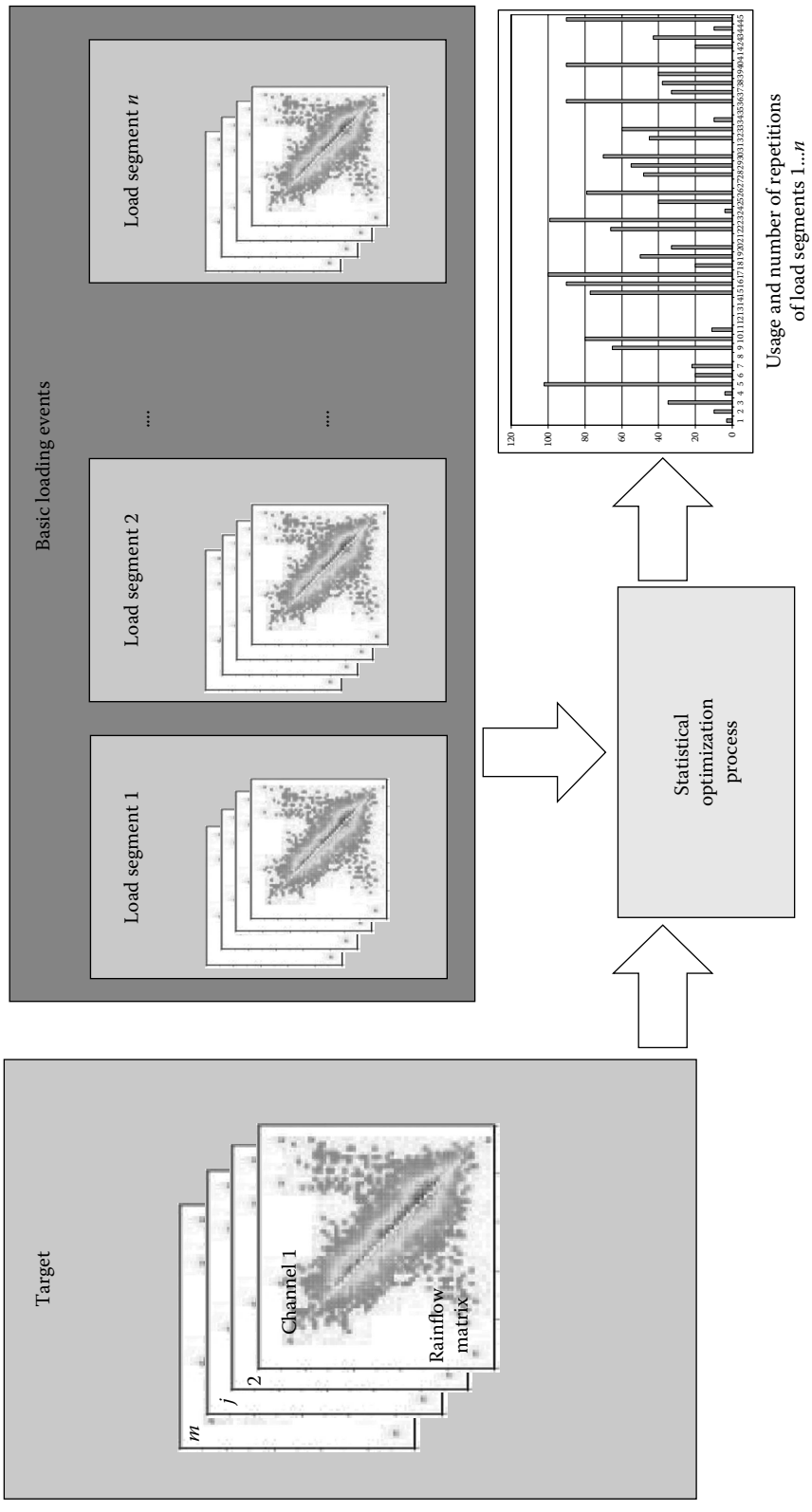


FIGURE 9.31 Approximation of target based on loading segments and their Rainflow matrices.

9.3.4.2 Frequency-Based Analyses

In principle, it is possible to generate the load-time sequences from the frequency region (by using the power spectral density together with other input parameters or, as the case may be, using assumptions) [64–67]. However, the corresponding procedures are only partially adequate in reproducing the extreme values of a signal.

The frequency content of a signal is vitally important for flexible and/or vibratory structures. Here, caution must be taken that the load amplitudes on the test rig are applied with the appropriate frequency spectrum. Typical application examples are axles or entire vehicles, but also shock absorbers, hydro bushings, or tuned mass dampers.

9.3.5 Standardized Load Sequence CARLOS TC

The previously described methods were used to compile the standardized load sequence CAR LOading Standard Trailer Coupling (CARLOS TC). This load sequence is used to design and test passenger car trailer couplings (TCs). Typically, apart from the visible hook (ball joint), the TC includes a construction consisting of a cross member and longitudinal struts, which incorporates the hook and transmits the trailer service loads into the chassis. These loads act on the coupling ball in the three spatial directions.

A considerable amount of measured data from TC service loadings together with the verification requirements based on the experience of several motor vehicle manufacturers was made available for deriving a three-channel load sequence. The D-value was used as a parameter for normalization, which is calculated from the mass of the pulling vehicle m_V and that of the trailer m_T (see Figure 9.32). By using the "Simultaneous Procedure" [59], the following objectives were realized: near service range-pair histograms and correlations between the force components as well as adapting the testing frequency, velocity, and acceleration of the test's control signals to the performance of the hydraulic actuators. The CARLOS TC load assumptions exist in the form of a three-channel load-time sequence that is subdivided into three modules for practical purposes. Figure 9.33 depicts the longitudinal force frequency distributions of the test modules M1, M2, and M3 with which the required overall spectrum is realized according to an iterative code.

Besides comparing the measured loads when appraising a prepared load sequence, it is also of interest to compare the local stresses and strains. Figures 9.34 and 9.35 [68] show this as an example of a relative comparison of local damage intensities caused by both loading with CARLOS TC and with loads, which were measured using a manufacturer's vehicle on a test track.

Finally, it should be pointed out that standardized load sequences can represent a benchmark for appraising load assumptions beyond direct applications for testing or designing.

9.4 ACCELERATED LIFE TESTING BY LABORATORY ROAD SIMULATION

Traditionally, the life testing of the vehicle and its major components has been performed on the road. Since long-term testing cannot be used within the vehicle development process, the life-cycle events have to be condensed significantly. Therefore, all car manufacturers developed so-called proving grounds and accelerated test strategies giving the capability to shorten the time for durability testing.

While road testing relies on the availability of complete vehicles, these tests are quiet costly. Consequently, the enhancements in building and controlling multiaxial test rig hardware that started in the 1960s were used to develop strategies for accelerated life testing by laboratory road simulation. Today, the laboratory-based durability approval is performed for many subsystems and will be even more important in the future because there are so many platform derivatives as well as niche products, which have to be developed within extremely short time scales [69].

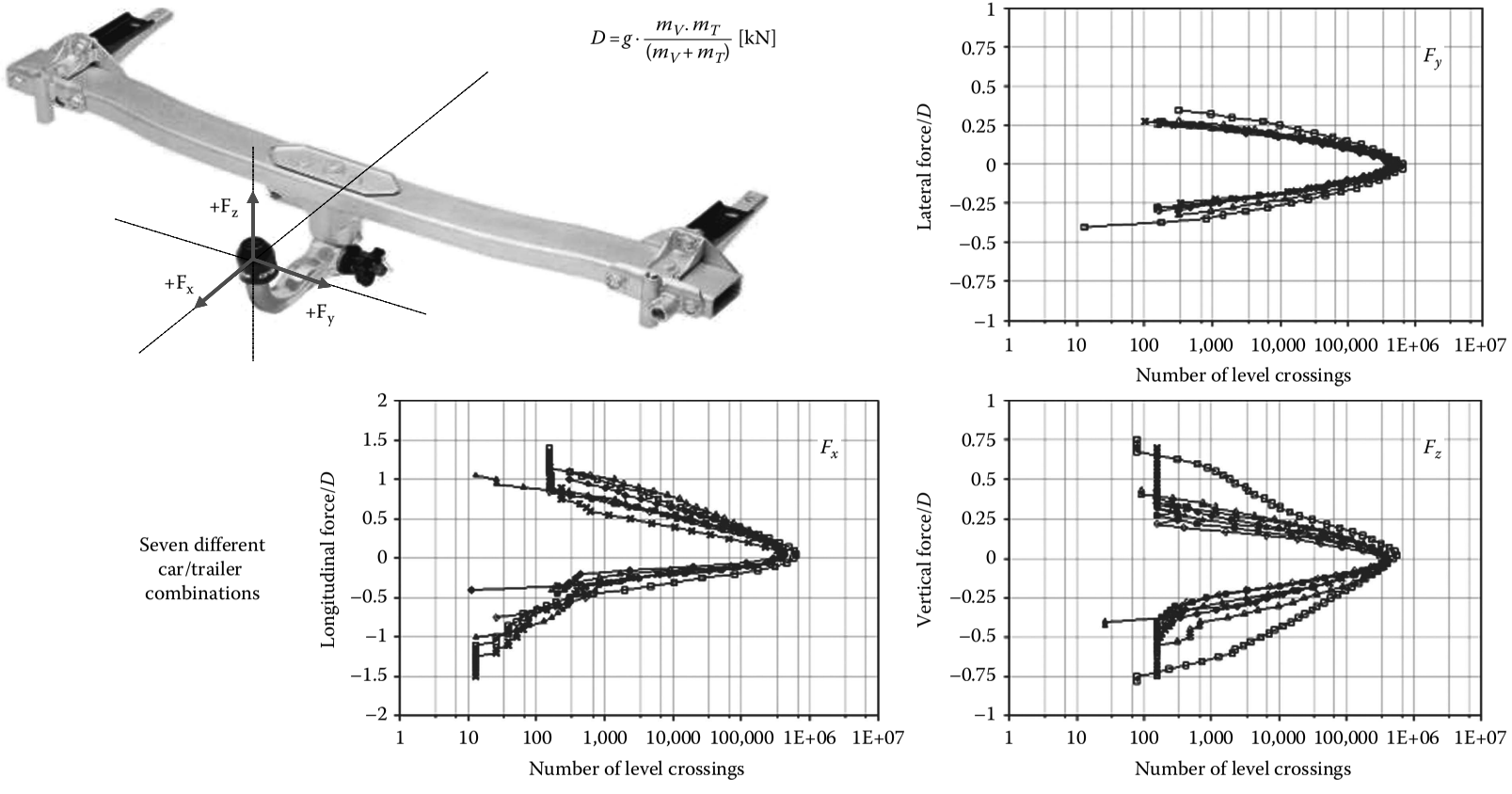


FIGURE 9.32 CARLOS TC: measured load spectra (normalized by *D*-value).

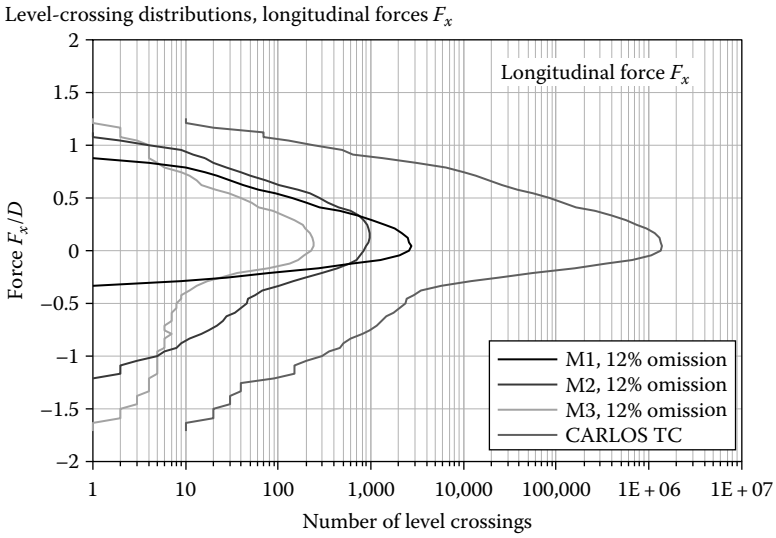


FIGURE 9.33 CARLOS TC: distributions of test modules M1, M2, M and whole test procedure 10*(5*(10*M1+M2)+M3).

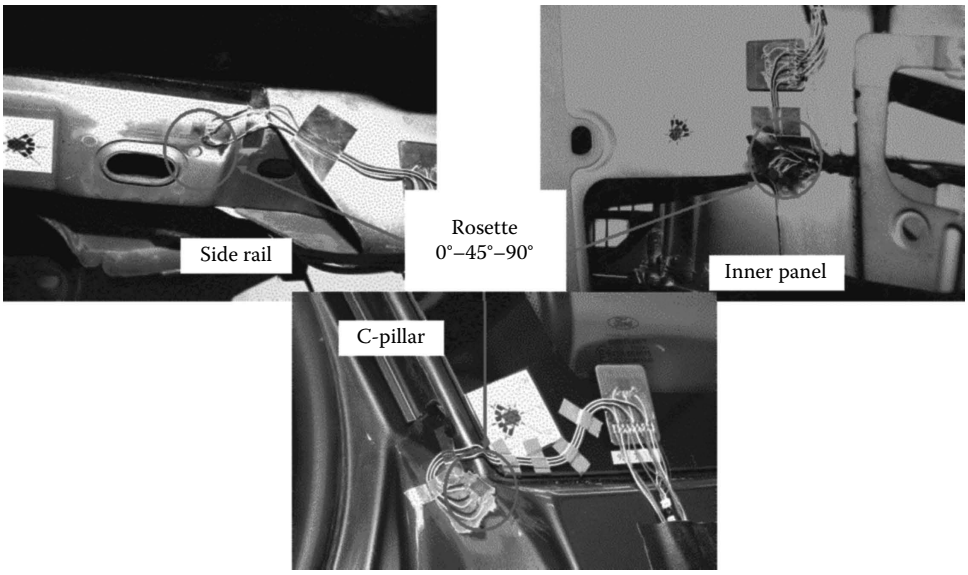


FIGURE 9.34 Measurement locations for correlation analysis on a Ford Focus (five doors).

9.4.1 Road Testing and Road Simulation

Usually, the final durability evaluation and approval is performed by road testing, which is carried out on a vast array of test beds, proving grounds, and public roads. These tests are a collection of worst driving and obstacle routes imaginable in customer operating scenarios—pot holes, rough cobblestones, speed bumps, non-surfaced roads, railway crossings and others. They are highly dynamic courses, on which chassis and driveline, vehicle body, and the running gear are tested [70].

Laboratory road simulation means the replication of conditions experienced by a vehicle in service—especially the simulation of motion and forces for the vehicle and its subsystems [71].

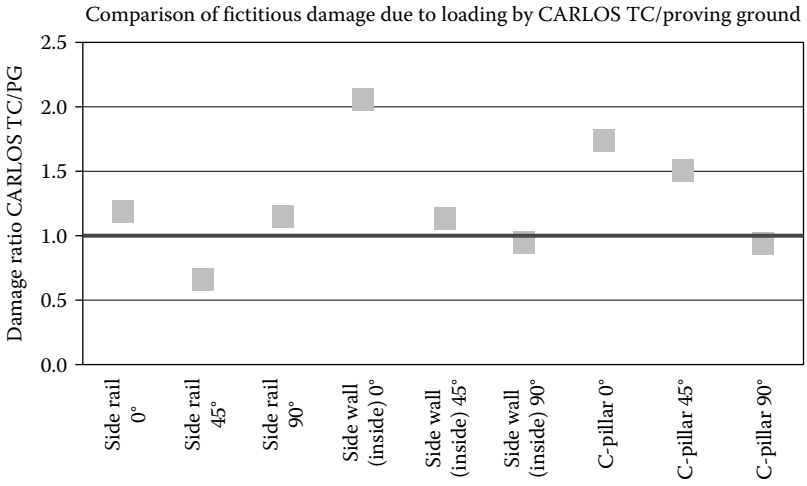


FIGURE 9.35 Correlation analysis between proving ground tests and CARLOS TC on a Ford Focus (five doors).

The advantage of service simulation is that these tests can be carried out in a controlled manner to enhance repeatability while offering savings in time and cost over road testing. Accelerated life testing is feasible, providing significant advantages for the vehicle development process [72].

Successful laboratory road simulation means that there is a complete understanding of the constraints and loading conditions and the practical limitations of the test rig hardware are known exactly.

9.4.2 Definition of System to Be Tested

The loading conditions are usually determined by measurements from real vehicle or workhorse tests on the road. Typically, wheel forces and motion, as the basic inputs for vehicle dynamics, are examined by using sensorized wheels or wheel force transducers. Hence, the basic excitation of the vehicle running on the ground is accessible, which helps to identify this portion of transient input data [73]. Actually the parts, which interface with the ground directly, are tires, wheels, and hubs—acting as unsprung rotating masses. The next subsystem is given by springs, dampers, and structural suspension components, such as struts and beams, having different kinematics compared to the rotating parts. Then the loading is transferred into the vehicle body and other major systems.

At this point there are a few main groups, which are different from each other in a couple of terms. Hence, the appropriate strategy for laboratory simulation clearly depends on the system to be examined. Since the suspension, the brakes, and the steering are safety critical subsystems, all of these need to be evaluated extremely carefully. Rotating components such as wheels or hubs are totally different to other suspension parts with regard to load kinematics; this has to be considered when choosing the test rig hardware for the laboratory simulation. Since the objective of the service simulation is to reproduce the damage of the operational conditions, subsystem constraints and loading have to be as close to reality as possible.

Therefore, laboratory road simulators are multiaxial loading systems to which the vehicle or one of its subsystems is attached.

9.4.2.1 Spindle-Coupled Road Simulators

For the interface between the ground and the vehicle, the road simulators often use spindle-coupled actuators for the loading operation [74]. A single corner may be used as an environment for testing and developing a quarter-car suspension system. By using two or four such corners, a complete



FIGURE 9.36 Road testing and laboratory road simulation.

suspension or even a full vehicle test rig can be configured [75]. Here, the vehicle can be tested as a fixed body having some ground coupling, or as a floating body simply supported by its inertia.

The four–post test rig provides the basic platform for carrying out various dynamic tests targeted at the vehicle chassis and body. More advanced test rig hardware enables up to 6 DOF per corner, which is the basis for a complete and sophisticated laboratory road simulation having vertical, lateral, and longitudinal translations, brake and drive torque, as well as steer and camber moment (Figure 9.36).

Accuracy issues may be caused by cross talk of the different channels, or some interference of the test rig loading devices and the vehicle suspension with regard to resonant frequencies.

9.4.2.2 Wheel and Hub Road Simulators

Since the load kinematics of wheels and hubs are totally different from suspension beams and struts, the use of spindle-coupled road simulators is not appropriate for durability testing of unsprung rotating components. Actually, these parts have cyclic loading sequences simply because the local stresses are given by the height of the external loads as well as the position relative to the load input area. Consequently, the design of a laboratory road simulator for wheels and hubs has to deal with the complete assembly including the tire as the element for transferring the loads into the subsystem.

The most complete laboratory road simulator for wheel durability testing is the biaxial wheel fatigue test machine (ZWARP) [76], which uses an inner drum as the moving ground (Figure 9.37). The wheel assembly is loaded by vertical and lateral actuators, which are capable of introducing service-like operational conditions such as straight driving on rough roads or heavy dynamic

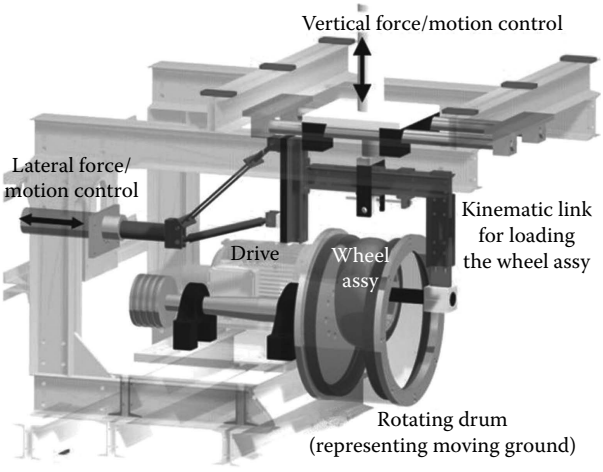


FIGURE 9.37 Biaxial wheel fatigue test machine (ZWARP).

cornering. Occasionally, torque loading can be introduced to examine the fatigue behavior of spoke-type wheels or performance issues such as unintentional wheel bolt loosening.

The so-called ZWARP technology is the leading international standard for wheel and hub laboratory road simulation [77]. Different to the biaxial simulation, there are more simplified durability procedures such as rotating bending or radial fatigue tests, which do not simulate service-like loading.

9.4.3 Load File Development

Vehicle dynamics implies loading, which is time-dependent. Hence, a special characteristic of using road simulators is the load file, which gives the force or motion time history for the individual channels. The load files can have standardized loading information, which is scalable for the specific vehicle, or they can be derived from individual road measurements.

The standardized load files include dedicated vehicle maneuvers such as straight driving, cornering, or braking, as well as information about load amplitudes and number of cycles [78]. They are generated semi-analytically by calculating the individual load histograms for all relevant vehicle maneuvers. Finally all load sequences are compiled and typically a block-type load file is created, which can be optimized very well for accelerated life testing.

While standardized loading information has its advantages, especially when road load data are not available—that is in an early stage of the vehicle development process or for a comparative benchmark of different vehicles—the dedicated load file accounts for all specific properties regarding vehicle and roads used. There is some significant amount of work to create such a load file having true time histories for the individual DOF (degree of freedom). By using techniques such as omission, extrapolation, or scaling, the basic load amplitude and phase correlations between the different load channels are kept, and the load file finally represents an accelerated test spectrum [79].

Since overstress acceleration of the test spectrum could cause severe impact on the damage mechanism by generating stresses beyond the structural yield limit, the more appropriate strategy is to apply usage rate acceleration.

Typically, the time scaling between the design life and the duration of the test is about 30 for suspension durability. This means the complete vehicle life cycle is performed within a period between 100 and 600 h. While the biaxial wheel fatigue testing using standardized load information and running at 100 kph takes only about 100 h, the range of frequency scaling for multiaxial suspension tests is limited by the test rig hardware as well as by the temperature shift of elastomeric bushings.

To control the different load and motion channels, digital controllers are used. The process of setting up highly accurate replications of multiaxial time histories is supported by road simulation software. These products create the test rig drive file typically by using an iteration strategy. Based on an identification process of the system's response, the initial load information is created and this is subsequently refined until the time histories are generated with high precision.

Since fatigue phenomena are the result of local strain time histories, it may cause inaccuracies when only forces and motion are controlled. Hence, there is often additional information from strain gauges at heavily stressed components, which has to be considered for the drive-file iteration. Highly sophisticated iteration strategies use virtual test laboratory support to reduce the effort for test rig hardware iteration. Therefore, a multi-body dynamics (MBD) representation of the test rig is used to identify improved initial loading information, which subsequently reduces the effort for hardware iteration.

9.4.4 Evaluation of Test Results

The repeatability of accelerated life testing by laboratory road simulation enables the test engineer to decide on the durability approval for the components tested. Since there is a huge amount of scatter with regard to durability performance—much more than dimensional tolerances only—an

TABLE 9.3
Multi-Tiered Failure Evaluation

Type	Object	Mode	Remark
I	Failure	Severe operational incident	Not allowed for safety relevant parts
II	Intervention	Unplanned occurrence, that requires interaction on maintenance level	
III	Event	Not related to failures	Engineering interest due to its frequency of occurrence

appropriate number of samples has to be tested. The exact number of samples depends on the strategy for reliability growth that is chosen. Within automotive development, there are usually a couple of samples available, which are used for censored test strategies. Therefore, all samples must achieve a specific test life target having none or minor fatigue cracks only [80].

Since complex products can have a large number of distinct failure modes, a certain degree of flexibility is called for in the failure definition. Hence, it is beneficial to have a multi-tiered failure evaluation process, which takes into account time to failure and its severity (Table 9.3).

Consequently, the final approval depends on the results of the multi-tiered failure evaluation process, which helps to prepare a rating for fatigue failure risk assessment.

REFERENCES

1. American Society for Metals (ASM). *Fatigue and Fracture. ASM Handbook*, Vol. 19. ASM International, Materials Park, OH (1996).
2. Bannantine, J.A., Comer, J.J., and Handrock, J.L. *Fundamentals of Metal Fatigue Analysis*. Prentice Hall, Englewood Cliffs, NJ (1990).
3. Buxbaum, O. Betriebsfestigkeit—Sichere und wirtschaftliche Bemessung schwingbruchgefährdeter Bauteile. Verlag Stahleisen GmbH Düsseldorf (1992), 2. erweiterte Auflage.
4. Haibach, E. Betriebsfestigkeit—Verfahren und Daten zur Bauteilberechnung. Springer-VDI Verlag Berlin (2006), 3. Auflage.
5. Radaj, D. and Vormwald, M. Ermüdungsfestigkeit, Grundlagen für Ingenieure. Springer-Verlag Heidelberg (2007), 3. Auflage.
6. Rice, R.C., Leis, B.N., Berns, H.D., Nelson, D.V., Lingenfleser, D., and Mitchell, M.R. *Fatigue Design Handbook* (SAE AE 10). Society of Automotive Engineers (SAE), Warrendale, PA (1988).
7. Sonsino, C.M. Dauerfestigkeit—Eine Fiktion. Konstruktion 57(4) (2005), 87–93.
8. Sonsino, C.M. Course of S-N curves especially in the high-cycle fatigue regime with regard to component design and safety. *International Journal of Fatigue* 29(12) (2007), 2246–2258.
9. Sonsino, C.M. and Grubisic, V. Requirements for operational fatigue strength of high quality cast components. *Materialwissenschaft und Werkstofftechnik* 27(8) (1996), 373–390.
10. Sonsino, C.M., Berg-Pollack, A., and Grubisic, V. Structural durability proof of automotive aluminium safety components—Present state of the art. Fatigue research and applications, SAE SP-1952, ISBN 978-0-7680-1597-3, SAE Paper 2005-01-0800/ (2005).
11. Waterhouse, R.B. Fretting fatigue. *International Materials Reviews* 37(2) (1992), 77–97.
12. Dixon, W.J. and Mood, A.M. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association* 43 (1948), 108–126.
13. Sonsino, C.M. Principles of variable amplitude fatigue design and testing. In: *Fatigue Testing and Analysis under Variable Amplitude Loading Conditions*. ASTM STP 1439 (2005), pp. 3–24, ASTM International, West Conshohocken, PA.
14. Sonsino, C.M. Fatigue testing under variable amplitude loading. *International Journal of Fatigue* 29(6) (2007), 1080–1089.
15. ASTM Standard E1049-85 (Reapproved 2005). *Standard Practices for Cycle Counting in Fatigue Analysis*. Beuth Verlag, Berlin, Germany (2005).
16. de Jonge, J.B. The analysis of load time histories by means of counting method. National Aerospace Laboratory (NLR); International Committee on Aeronautical Fatigue (ICAF) (NLR MP 82039 U) (1982).

17. Sonsino, C.M. Werkstoffauswahl für schlagartig und zyklisch belastete metallische Bauteile. *Materialwissenschaft und Werkstofftechnik* 32(3) (2001), 221–230.
18. ASTM Standard E606-92. Standard practice for strain-controlled fatigue testing. In: *Annual Book of ASTM Standards*, Vol. 03.01. ASTM (1997), pp. 523–537, ASTM International, West Conshohocken, PA.
19. Sonsino, C.M., Zenner, H., Yousefi-Hashtyani, F., and Küppers, M. Assessments of components under multiaxial service loading—Calculations and experiments. In: *Centre Technique des Industries Méchaniques (CETIM)*. Fatigue Design (2005), 11p.
20. Sonsino, C.M. Zur Bewertung des Schwingfestigkeitsverhaltens von Bauteilen mit Hilfe örtlicher Beanspruchung. *Konstruktion* 45(1) (1993), 25–33.
21. Sonsino, C.M., Kaufmann, H., and Grubisic, V. Übertragbarkeit von Werkstoffkennwerten am Beispiel eines betriebsfest auszulegenden geschmiedeten Nutzfahrzeug-Achsschenkels. *Konstruktion* 47(7/8) (1995), 222–232.
22. Sonsino, C.M. Light-weight design chances using high-strength steels. *Materialwissenschaft und Werkstofftechnik* 38(1) (2007), 9–22.
23. Kloos, K.H., Fuchsbauer, B., Magin, W., and Zankov, D. Übertragbarkeit von Probestab-Schwingfestigkeitseigenschaften auf Bauteile. In: *VDI-Berichte* (1979), Nr. 354, pp. 59–72.
24. Krae, Ch., Heuler, P., and Heckel, K. Consideration of size effects for assessment of notch fatigue life. International Committee on Aeronautical Fatigue (ICAF 1990) (ICAF Doc. 1973).
25. Bomas, H., Linkewitz, T., and Mayr, P. The weakest link concept—A probabilistic approach to the fatigue limit of components. In: *MATERIALS WEEK 2000—Proceedings*; editor and organiser: Werkstoffwoche-Partnerschaft. Frankfurt, Germany (September 25–28, 2000), URL: www.materials-week.org/proceedings
26. Sonsino, C.M., Kaufmann, H., and Grubisic, V. Transferability of material data for the example of a randomly loaded forged truck stub axle. Recent developments in fatigue technology, SAE PT-67, ISBN 978-0-7680-0037-5, SAE Paper 970708 (1997), S. 649–670.
27. Rechnerischer Festigkeitsnachweis für Maschinenbauteile aus Stahl, Eisenguss- und Aluminiumwerkstoffen. Forschungskuratorium Maschinenbau, Frankfurt, 4. erweiterte Ausgabe (2002).
28. Hobbacher, A. Recommendations for fatigue design of welded joints and components. *International Institute of Welding*, doc. XIII-1965-03/XV-1127-03 (2003).
29. Bruchmechanischer Festigkeitsnachweis. Forschungskuratorium Maschinenbau, Frankfurt, 1. Ausgabe (2001).
30. Nieslony, A. and Sonsino, C.M. Comparison of some selected multiaxial fatigue assessment criteria. Fraunhofer Institute for Structural Durability and System Reliability LBF, Darmstadt. Report No. FB-234 (2008).
31. Sonsino, C.M. Multiaxial fatigue assessment of welded joints—Recommendations for design codes. *International Journal of Fatigue* 31(1) (2009), 173–187.
32. Döring, R. Zum Deformations- und Schädigungsverhalten metallischer Werkstoffe unter mehrachsig nichtproportionalen zyklischen Beanspruchung. Dissertation, TU Darmstadt (2006).
33. Hoffmeyer, J. Anrisslebensdauervorhersage bei mehrachsiger Beanspruchung auf Basis des Kurzrisskonzepts. Dissertation, TU Darmstadt, Darmstadt, Germany (2005).
34. Socie, D.F. and Marquis, G.B. *Multiaxial Fatigue*. Society of Automotive Engineers, Warrendale, PA (2000).
35. Sonsino, C.M., Zenner, H., Yousefi-Hashtyani, F., and Küppers, M. Present limitations in the assessment of components under multiaxial service loading. In: *Proceedings of 7th ICBMFF*. DVM Berlin, Germany (2004), pp. 17–26.
36. Eulitz, K.-G. and Kotte, K.L. Damage accumulation—Limitations and perspectives for fatigue life assessment. In: *MATERIALS WEEK 2000—Proceedings*; editor and organiser: Werkstoffwoche-Partnerschaft. Frankfurt, Germany (September 25–28, 2000), URL: www.materialsweek.org/proceedings
37. Sonsino, C.M. *Fatigue Design Concepts for P/M Parts and Required Material Data—An Overview*. Metal Powder Industries Federation, Princeton, NJ (2003).
38. Filippini, M. and Dieterich, K. An approximate formula for calculating the probability of failure. Fraunhofer-Institut für Betriebsfestigkeit LBF, Darmstadt. Technische Mitteilungen TM-Nr. 111 (1997).
39. Sonsino, C.M. Versuchszeitverkürzung in der Betriebsfestigkeitsprüfung—Methoden und Kompromisse. *Materialprüfung* 45(4) (2003), 133–144.
40. Klätschke, H. and Schütz, D. Das Simultanverfahren zur Extrapolation und Raffung von mehraxialen Belastungs-Zeitfunktionen für Schwingfestigkeitsversuche. *Materialwissenschaft und Werkstofftechnik* 26 (1995), Nr. 8, 404–415.

41. Klätschke, H. and Heuler, P. Generation and use of standardised load spectra and load-time histories. *International Journal of Fatigue* 27(8) (2005), 974–990.
42. Kieninger, M., Rupp, A., and Gerlach, T. New wheel force transducer for the investigation of multiaxial loads on rail bound vehicles in the urban traffic, 37. Tagung Moderne Schienenfahrzeuge Technische Univeristät Graz, 15. bis 18 (April 2007), S. 102–109.
43. Hobbacher, A. Recommendations for Fatigue Design of Welded Joints and Components. International Institute of Welding, doc. XIII-1965-03/XV-1127-03. Paris, France, 2005.
44. Gaßner, E. Festigkeitsversuche mit wiederholter Beanspruchung im Flugzeugbau. Luftwissen (1939), Nr. 6, S. 61–64.
45. Svenson, O. Beanspruchung und Lastkollektiv am Fahrwerk von Kraftfahrzeugen. *Automobiltechnische Zeitschrift* 65(11) (1963), 334–337.
46. Buxbaum, O. Statistische Zählverfahren als Bindeglied zwischen Beanspruchungsmessung und Betriebsfestigkeitsversuch. Fraunhofer-Institut für Betriebsfestigkeit (LBF), Darmstadt, Bericht Nr. TB-65 (1966).
47. Haibach, E. Beurteilung der Zuverlässigkeit schwingbeanspruchter Bauteile. *Luftfahrttechnik-Raumfahrttechnik* 13(8) (1967), 188–193.
48. Grubisic, V. Bemessung und Prüfung von Fahrzeug-Rädern. *Automobiltechnische Zeitschrift* 75(1) (1973), 9–18 and (7), 252–258.
49. de Jonge, J.B. Counting methods for the analysis of load time histories. NLR Memorandum SB-80-106 U (1980).
50. Buxbaum, O. Betriebsfestigkeit, sichere und wirtschaftliche Bemessung schwingbruchgefährdeter Bauteile. Verlag Stahleisen mbH, Düsseldorf, 2. Auflage (1992).
51. Schütz, D., Klätschke, H., and Heuler, P. Standardisierte mehraxiale Lastabläufe für Bauteile von Pkw-Radaufhängungen—CAR LOading Standard CARLOS multi. Fraunhofer-Institut für Betriebsfestigkeit (LBF), Darmstadt, Bericht Nr. FB-201 (1994).
52. Westermann-Friedrich, A. and Zenner, H. Zählverfahren zur Bildung von Kollektiven aus Zeitfunktionen – Vergleich der verschiedenen Verfahren und Beispiele. Institut für Hüttenmaschinen und Maschinelle Anlagentechnik der TU Clausthal, FVA-Merkblatt Nr. 0/14 (1988).
53. Fischer, G. Experimenteller Nachweis an ausgewählten Bauteilen des Schienenfahrzeugbaus. DVM-Weiterbildungsseminar, Lebensdauer und Betriebsfestigkeits—nachweis, Koblenz (Oktober 2001), DVM Berlin.
54. Horst, N., Schäfer, U., and Schmidt, R. Ermittlung von statistisch abgesicherten Kunden-Lastkollektiven für Personenkraftwagen. In: *Fahrwerke und Betriebsfestigkeit, DVM Bericht* 129 (2002), 81–91.
55. Johannesson, P. Extrapolation of load histories and spectra. *Fatigue & Fracture of Engineering Materials & Structures* 29(3) (2006), 201–207.
56. Dressler, K., Gründer, B., Hack, M., and Köttgen, V.B. Extrapolation of Rainflow matrices, SAE Technical Paper 960569 (1996).
57. Johannesson, P. and Thomas, J.J. *Extrapolation of Rainflow Matrices*. Extremes 4:3, Kluwer Academic Publishers, Dordrecht, the Netherlands (2002), pp. 241–262.
58. Clormann, U.H. and Seeger, T. Rainflow-HCM. Ein Hysteresisschleifen-Zählalgorithmus auf werkstoffmechanischer Grundlage, Programmbeschreibung. Bericht FF-22, Fachgebiet Werkstoffmechanik der TH Darmstadt (1985).
59. Klätschke, H. and Schütz, D. Das Simultanverfahren zur Extrapolation und Raffung von mehraxialen Belastungs-Zeitfunktionen für Schwingfestigkeitsversuche. *Materialwissenschaft und Werkstofftechnik* 26(8) (1995), 404–415.
60. Flade, D., Helm, B., and Lücker, E. Simulation von schädigungsäquivalenten Beanspruchungen an Elastomerbauteilen (Simsabel). Final report of a research project supervised by AK Elastomerlager, LBF-Bericht Nr. 112100 (2003), unpublished.
61. Helm, B., Bruder, T., and Flade, D. Betriebsfestigkeitsnachweis von Elastomerbauteilen. Konstruktion 56 (2004), IW 14–IW 16.
62. Zwittian, T. Strategie zur Prüfsignalkürzung für multifunktionale Elastomerlager. DVM-Workshop, Prüfmethodik für Betriebsfestigkeitsversuche in der Fahrzeugindustrie, Darmstadt (January 31/February 1, 2006).
63. Gründer, B., Speckert, M., and Pompetzki, M. *Design of Durability Sequences Based on Rainflow Matrix Optimization*. Society of Automotive Engineers, Warrendale, PA (1998).
64. Bishop, N.W.M. Spectral methods for estimating the integrity of structural components subjected to random loading. In: *Handbook of Fatigue Crack Propagation in Metallic Structures*. Elsevier Science B.V., Amsterdam, the Netherlands (1994), pp. 1685–1720.

65. Petrucci, G. and Zuccarello, B. On the estimation of the fatigue cycle distribution from spectral density data. *Journal of Mechanical Engineering Science* C213 (1999), 819–831.
66. Petrucci, G. and Zuccarello, B. Fatigue life prediction under wide band random loading. *Fatigue & Fracture of Engineering Materials & Structures* 27 (2004), 1183–1195.
67. Benasciutti, D. and Tovo, R. Spectral methods for lifetime prediction under wide-band stationary processes. *International Journal of Fatigue* 27 (2005), 867–877.
68. Bruder, T., Klätschke, H., Sigwart, A., and Riehle, J. Leichtbau und Betriebsfestigkeit durch realitätsnahe Lastannahmen am Beispiel von PKW-Anhängevorrichtungen. In: *Materialprüfung* 47(7–8) (2005), 429–436.
69. Ledesma, R., Jenaway, L., Wang, Y., and Shih, S. Development of accelerated durability tests for commercial vehicle suspension components, SAE Paper 2005-01-3565 (2005).
70. Haq, S., Temkin, M., Black, L., and Bammel, P. Vehicle road simulation testing, correlation and variability, SAE Paper 2005-01-0856 (2005).
71. Jeong, G.-S., Moon, H.-S., and Sung, D.-U. The development of lab-simulation test to accelerate the durability validation of engine mounting and wiring, SAE Paper 2003-01-0949 (2003).
72. Heim, R., Fischer, G., and Sonsino, C.M. Early stage rig testing for durability approval, SAE Paper 2006-01-0116 (2006).
73. Rupp, A., Grubisic, V., and Neugebauer, J. Development of a multi-component wheel force transducer—A tool to support vehicle design and validation, Warrendale, PA, SAE Technical Papers 930258 (1993).
74. White, K.J. The road simulator—A practical laboratory approach, IMechE C110/85, 1985.
75. Toshiaki, N. Durability test with a road simulator for body and chassis parts durability development, *Journal of the Society of Automotive Engineers of Japan* 59(7) (2005), 89–94.
76. Grubisic, V. and Fischer, G. Biaxial wheel/hub test facility. In: *Proceedings of the 2nd International User Meeting*, LBF, Darmstadt, Germany (1996).
77. SAE International, Wheel Standards Committee. Surface Vehicle Recommended Practice—Biaxial Wheel Fatigue Test, SAE J2562_200511 (2005).
78. Wessling, U., Fischer, G., and Grubisic, V. Methodik zur effektiven Dimensionierung und Bewertung der Betriebsfestigkeit von Fahrwerkskomponenten, 23. Tagung des DVM-AK Betriebsfestigkeit, Köln, Hrsg. Deutscher Verband für Materialforschung und -prüfung, Berlin, DVM-Bericht, 123 (1997).
79. Heim, R., Krause, I., and Weingaertner, S. Runflat—Technology and its impact on design and durability of wheels, SAE Paper 2007-01-1532 (2007).
80. Grubisic, V. and Fischer, G. Methodology for effective design evaluation and durability approval of car suspension components, SAE Paper published in SP-1223 (1997).

10 Reliability Assessment of Mechatronic Devices in Vehicles

Bernd Bertsche, Jochen Gäng, Holger Hanselka, Soong-Oh Han, Jürgen Nuffer, and Kai Wolf

CONTENTS

10.1 Introduction 308
10.2 General Aspects of Reliability 308
10.2.1 Quantitative Reliability 308
10.2.1.1 Histogram and Density Function $f(t)$ 310
10.2.1.2 Distribution Function or Failure Probability $F(t)$ 311
10.2.1.3 Survival Probability or Reliability $R(t)$ 311
10.2.1.4 Failure Rate $\lambda(t)$ 311
10.2.1.5 Durability Distributions 312
10.2.1.6 Weibull Distribution 312
10.2.1.7 Boolean Theory 313
10.2.1.8 Fault Tree Analysis 314
10.2.2 Qualitative Method—Failure Modes and Effects Analysis 314
10.3 Reliability Assessment in Early Development Phases 315
10.3.1 Development Process and the V-Model 315
10.3.2 Reliability Methodology for Mechatronic Systems in Early Development Phases 316
10.3.2.1 Identification of Product Function/Malfunction 316
10.3.2.2 Detailed System Design 319
10.3.2.3 Data Collection 319
10.3.2.4 Reliability Modeling 319
10.3.2.5 Calculation 319
10.4 Numerical Sensitivity Analysis and Reliability Investigation of an Active Noise Reduction System 320
10.4.1 Performance Evaluation 320
10.4.2 Sensitivity Analysis 321
10.4.3 Factors Influencing Reliability 323
10.4.4 Lifetime Estimation 324
10.5 Conclusion 324
References 325

10.1 INTRODUCTION

Over the last two decades in particular, devices combining mechanical, electrical, and software functionalities, so-called mechatronic systems, became more and more important within all branches of the modern mass market. The reliability assessment of mechatronic systems is regarded to be one of the key issues in their acceptance. The present chapter aims to give an overview on the state of the art and some recent developments in this field.

In the first section of this chapter, a brief introduction to the theoretical background of reliability engineering is given (Section 10.1). Here, general aspects of reliability are addressed, such as failure probabilities represented by statistical distributions for quantitative analysis. Also, a well-established procedure for qualitative reliability analysis is introduced.

In the second section, some aspects of the reliability assessment of mechatronic systems in early development phases are presented (Section 10.2). Information about the design details of a system may not be available in an early stage. It is demonstrated how a reliability estimate can be obtained from existing data.

With increasing complexity of the systems, experimental costs become exceedingly large. For this reason, great efforts are taken to replace traditional test methods by computational simulation tools which have to be able to reproduce the complex system performance and also account for various failure and fatigue scenarios. The third section of this chapter presents an example of a qualitative numerical reliability investigation of a complex prototype automotive noise reduction system (Section 10.3).

Here, dominant failure modes are identified via numerical modeling and sensitivity analysis.

10.2 GENERAL ASPECTS OF RELIABILITY

The development of modern products is currently confronted with increasing functional requirements, higher complexity, interconnectedness of hardware, software, and sensor technology as well as low development and product costs. What significance does reliability have and what is even meant by this concept? The concept of reliability is connected to the functional capability of a product [1]. According to the VDI Guideline 4001 [2], reliability is defined as follows:

Reliability is the probability that a product does not fail during a specified time interval under given functional and environmental conditions.

To reach a high level of customer satisfaction, the reliability of the system must be regarded as the prime issue during the entire product development process. With regard to this, suitable organizational and substantive measures must be implemented [1]. It is important that all groups along the chain of development are included: errors occur at all phases. The available methodological tools should be appropriately chosen during the entire product lifecycle and consistently implemented in a coordinated way (Figure 10.1).

Reliability analyses can be subdivided into qualitative and quantitative approaches [1,3]. As the aim of qualitative analysis is a systematic investigation of failure, the aim of quantitative analysis is to calculate the reliability. Figure 10.2 shows different qualitative and quantitative reliability methods [1].

In the following, a few essential basic principles of reliability engineering are presented. On the one hand, fundamental terms for the calculation of the reliability as well as different distributions are discussed. On the other hand, the most established qualitative method—the failure mode and effects analysis (FMEA)—is presented.

10.2.1 Quantitative Reliability

To describe failure behavior, the following statistical functions can be used: Probability density function $f(t)$, distribution or failure probability function $F(t)$, survival probability or reliability function $R(t)$, and the failure rate $\lambda(t)$.

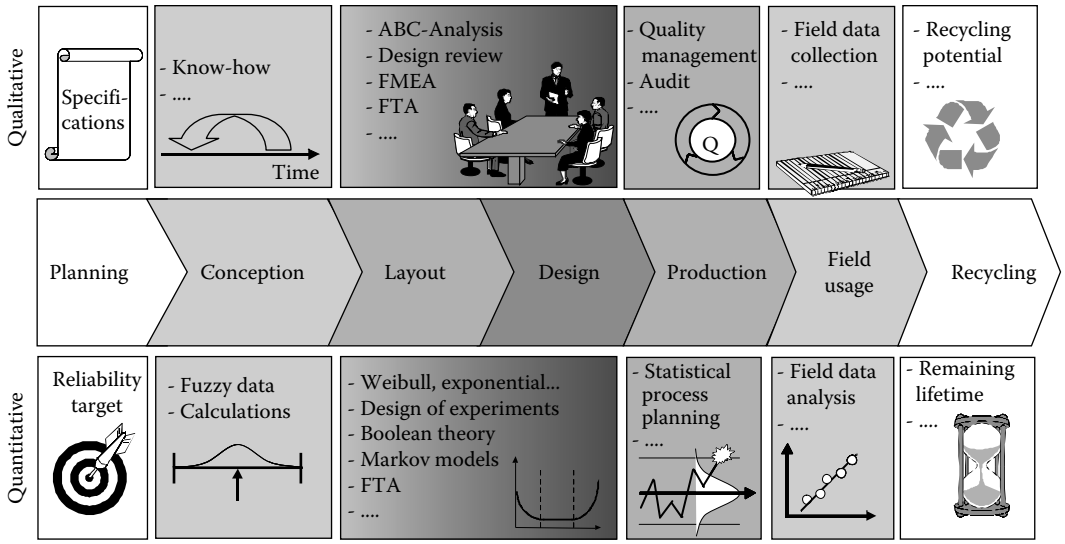


FIGURE 10.1 Qualitative and quantitative reliability methods in different phases of the product life cycle. (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

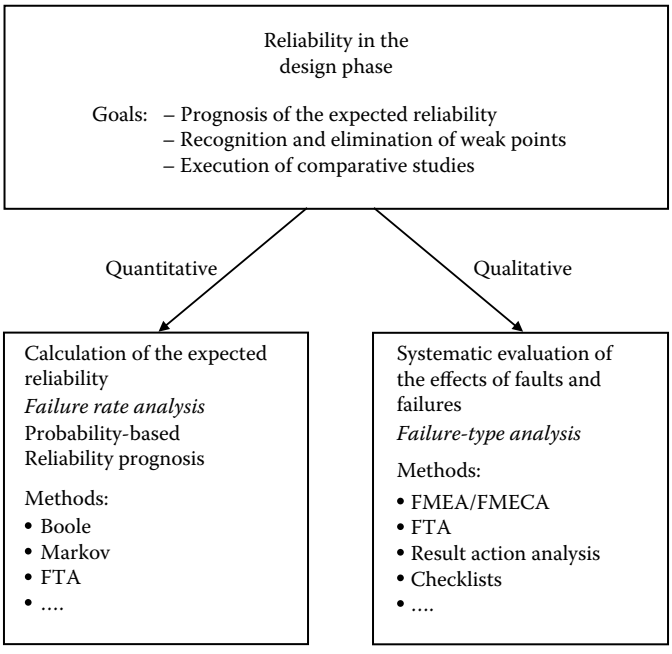


FIGURE 10.2 Goals as well as differentiation of quantitative and qualitative reliability methods in the design phase. (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

10.2.1.1 Histogram and Density Function *f*(*t*)

A histogram of the failure frequency provides the simplest option for graphically representing failure behavior. Figure 10.3 shows a histogram which is based on a Wöhler test for the frequency of a defined load.

The times of failure or rather the load cycles at which the component fails occur in a certain period purely randomly. In order to assign these times of failure, the number n, the abscissa, on which the times of failure are plotted, is divided into different classes n_k. In these classes, the number of failures is determined. Since the number of classes is not easily defined—too few classes result in a loss of information and gaps can occur with too many classes—a rough approximation or estimated value for the number of classes n_k is given by

$$n_k \approx \sqrt{n} \tag{10.1}$$

Instead of a histogram, the failure behavior can also be described by the empirical density function $f^*(t)$. Here, the midpoints of the tops of each column in the histogram are connected to each other. The underlying ideal density function $f(t)$ is to be found when the number n of the tested components is increased. In the limit as $n \to \infty$, the histogram's outline approaches a smooth continuous curve. The limit as $n \to \infty$ implies that all components of a very large population were tested and thus the failure behavior was precisely determined (Figure 10.4).

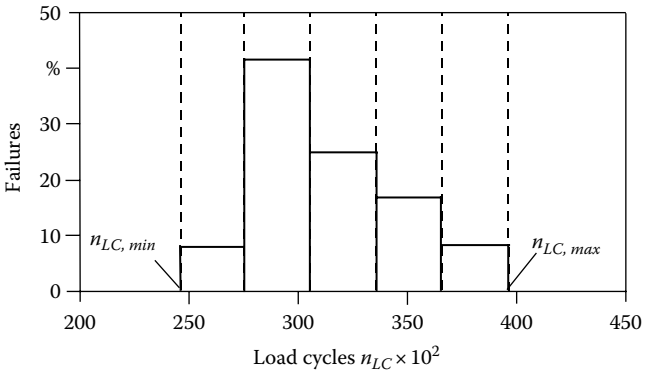


FIGURE 10.3 Histogram for the failure frequency depending on the number of load cycles. (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

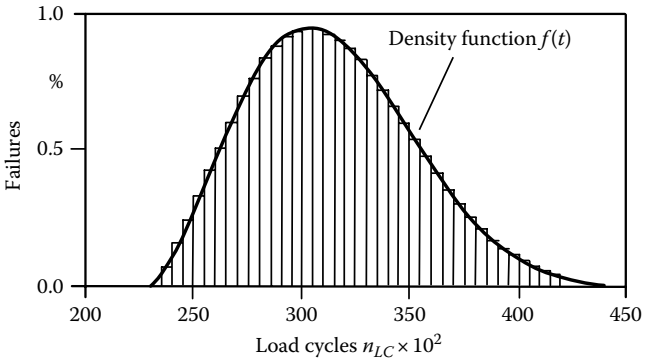


FIGURE 10.4 Histogram of failure frequencies and density function $f(t)$. (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

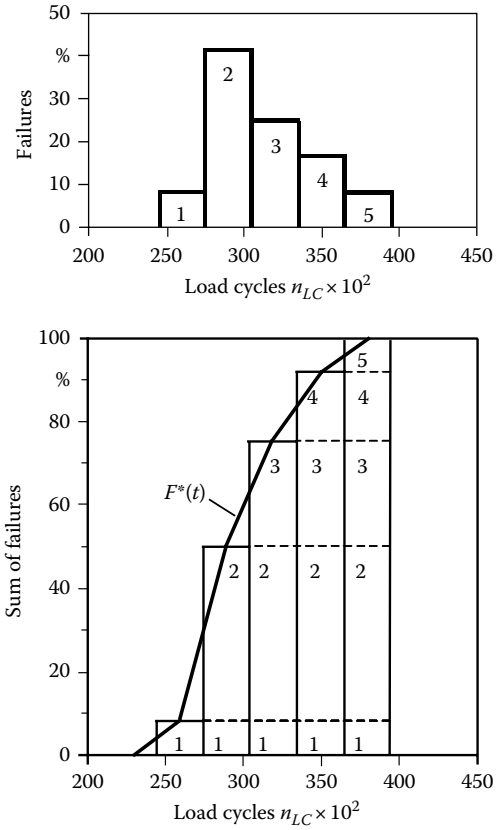


FIGURE 10.5 Histogram and distribution function. (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

10.2.1.2 Distribution Function or Failure Probability *F*(*t*)

In order to obtain the distribution function, that is, how many components have failed at a particular time, the observed failures from the histogram are added up (Figure 10.5).

10.2.1.3 Survival Probability or Reliability *R*(*t*)

The survival probability $R(t)$ and the distribution function are interdependent as follows:

$$R(t) = 1 - F(t) \tag{10.2}$$

Since at $t = 0$ failure has not yet occurred, the survival probability $R(t)$ always begins at $R(0) = 1$. In reliability theory, the function is referred to as the reliability $R(t)$.

10.2.1.4 Failure Rate λ(*t*)

Describing the failure behavior with the failure rate $\lambda(t)$, the failures at a time t or in a class i are not related to the sum of failures, as with the relative frequency, but to the sum of the surviving units. This implies that the failure rate $\lambda(t)$ is the quotient of the density function $f(t)$ and the survival probability $R(t)$:

$$\lambda(t) = \frac{f(t)}{R(t)} \tag{10.3}$$

The failure rate is frequently used to describe early, random, and wear-out failures. An attempt is made to register the overall failure behavior of a component or a machine. This is done using the "bathtub curve" (Figure 10.6) [1,3].

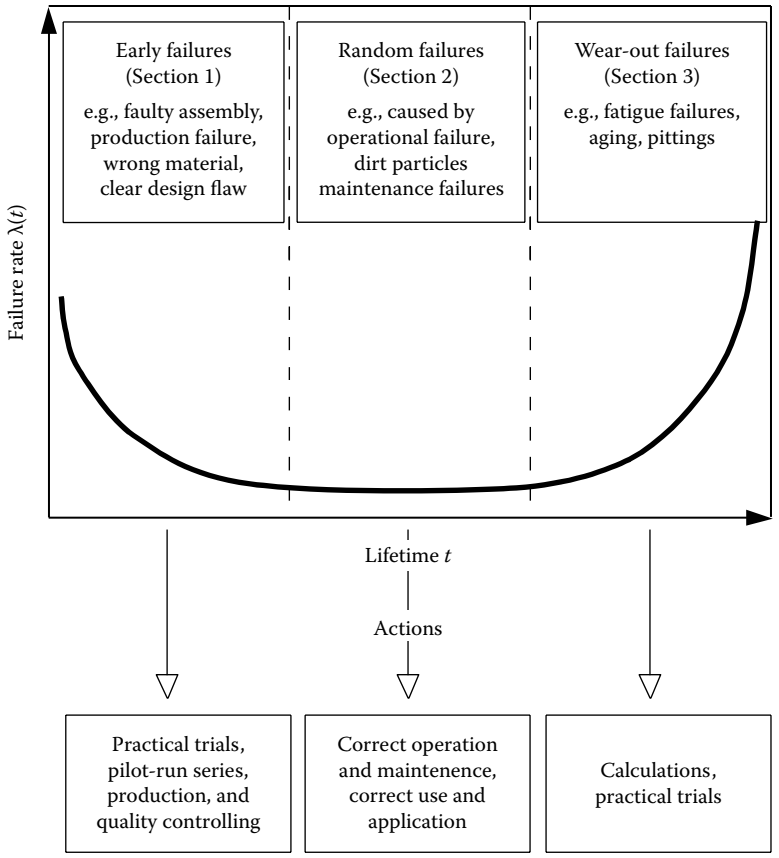


FIGURE 10.6 Three sections of the failure rate (bathtub curve). (From Bertsche, B., *Reliability in Automotive and Mechanical Engineering*, Springer, Berlin, Germany, 2008.)

Here, three sections are distinguished. In the first section, early failures are to be found, which result from faulty assembly, production failures, or wrong material. The second section represents random failures caused by operational failure, dirt particles, or maintenance failures. The third section describes wear and fatigue failures, which occur through fatigue fracture, aging, or pitting.

10.2.1.5 Durability Distributions

Durability distributions mathematically describe the progress of failure behavior. The most well-known "durability distribution" is the Gaussian or normal distribution [1,3,4] which is nevertheless only very rarely used in reliability theory. The exponential distribution is frequently employed in electrical engineering and the Weibull distribution is commonly used in mechanical engineering [1,3,4]. The log-normal distribution is occasionally employed in materials science and mechanical engineering [1].

10.2.1.6 Weibull Distribution

Very different failure behaviors can be described with the Weibull distribution (Figure 10.7a and b). This is most clearly demonstrated by representing the density function with the help of the Weibull function. Subject to one of the two distribution parameters—the shape parameter b—the density function changes considerably. For small values of b ($b < 1$), failures occur which are typical of early failures. A constant failure rate λ is found with $b = 1$. For $b > 1$, the density function always begins at $f(t) = 0$. After increasing, the fatigue life reaches a maximum and finally falls gradually to low values. For increasing values of b, the density function's maximum is translated further to the right until the value $b = 3.5$ is reached, where the density function becomes symmetrical.

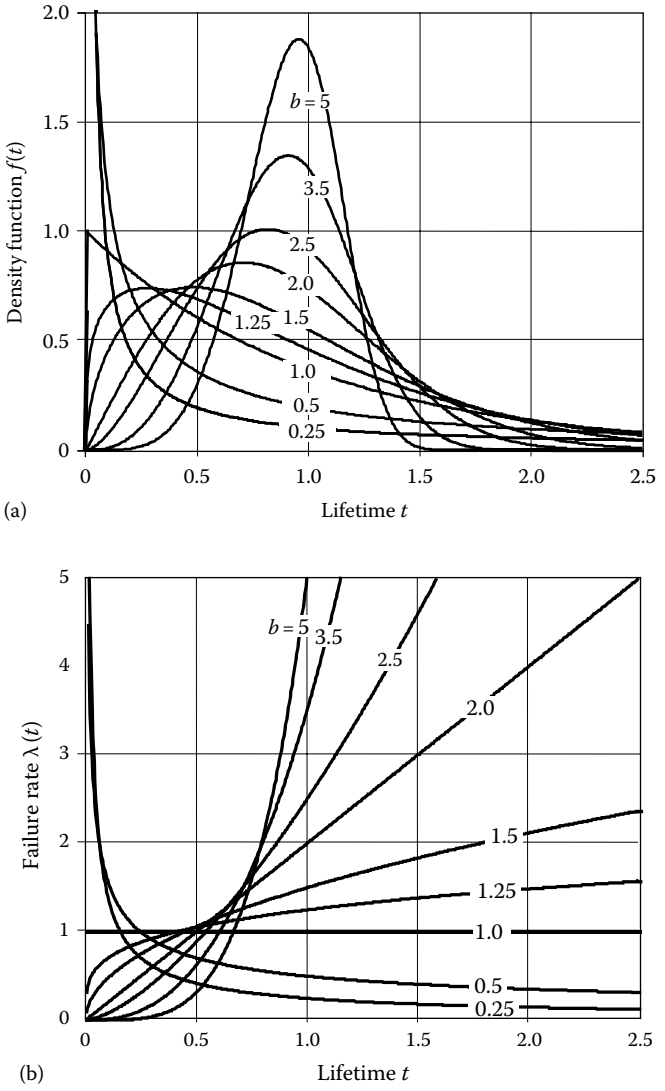


FIGURE 10.7 (a) Density function of the Weibull distribution. (b) Failure rate of the Weibull distribution.

Apart from the graphical considerations, it is possible to describe the Weibull distribution mathematically (Table 10.1).

[1,3–7] gives the other durability distributions as well as a more detailed insight of the reliability analysis.

10.2.1.7 Boolean Theory

Boolean theory represents the simplest system theory. Starting from the component failure behavior, the failure behavior of the system can be computed using the Boolean system theory. As described in Section 10.2.1.5, the failure behavior of individual components can be represented [1]. To implement the Boolean theory, the following prerequisites must be satisfied:

- The system is "non-repairable," that is, the first system failure ends the system's life.
- The system's elements can only assume two states: "operative" or "failed."
- The system's elements are "independent," that is the failure of one component is not influenced by the failure of another component.

TABLE 10.1
Formulae for the Weibull Distribution with the Shape Parameter *b* and the Scale Parameter *T*

$R(t) = e^{-(t/T)^b}$	Reliability function
$\lambda(t) = \frac{f(t)}{R(t)} = \frac{b}{T} \cdot \left(\frac{t}{T}\right)^{b-1}$	Failure rate
$F(t) = 1 - e^{-(t/T)^b}$	Failure function
$f(t) = \frac{dF(t)}{dt} = \frac{b}{T} \cdot \left(\frac{t}{T}\right)^{b-1} e^{-(t/T)^b}$	Failure density function

Using the system's elements, a reliability circuit diagram can be drawn up, from which the reliability configuration of a system is discernible [1]. At the same time, the reliability circuit diagram shows how the failure of components affects the overall system. The system is then operative if and only if a connection between the input and output in the reliability circuit diagram exists and on which all the depicted components are intact. The failure of an arbitrary component in a serial configuration leads to the failure of the overall system. In a parallel configuration, the system does not fail until all components have failed.

10.2.1.8 Fault Tree Analysis

Fault tree analysis (FTA) is a deductive method that serves to illustrate the system's function and to quantify the system's reliability. The method can be used both as a diagnostic and as a development tool. In this way, potential failures in a system can be identified and alternative designs can be evaluated. One of the advantages underlying the FTA is that the method yields both qualitative and quantitative results. FTA is based on Boolean theory.

10.2.2 Qualitative Method—Failure Modes and Effects Analysis

Failure modes and effects analysis (FMEA) is the most commonly and most frequently implemented method in reliability engineering. FMEA was developed in the United States by NASA [1] in the mid-1960s for the Apollo project. It was subsequently used as a general method in the aerospace industry. In the meantime, FMEA represents a standard method in the automobile industry both in Europe and also in Japan and the North America Free Trade Agreement (NAFTA). German standards can be found in [8], international standards in [9,10]. As extension, the failure mode effects and criticality analysis (FMECA) can be used. Compared with the FMEA, the FMECA includes a criticality analysis [11].

FMEA is a systematic method whose fundamental idea is to identify all possible failure modes for arbitrary systems, subsystems, and components. Simultaneously, the possible failure effects and causes are identified. Finally, risks are evaluated and optimization measures are specified. The aim of the method is to identify a product's risks or weak spots as soon as possible in order to carry out timely improvements.

The implementation of FMEA is carried out in interdisciplinary groups or so-called FMEA teams. This is expedient since only then it is guaranteed that all parties involved in the manufacturing chain can satisfactorily participate.

The five stages of the FMEA system according to [8] are depicted in Figure 10.8 and explained as follows:

Step 1: System elements and system structure. Definition of the system's configuration (system, assemblies, components)

Step 2: Functions and function structure. Definition of the system's, the assemblies', and the component's functions, e.g., by means of coordination, previous FMEA, requirements, specifications.

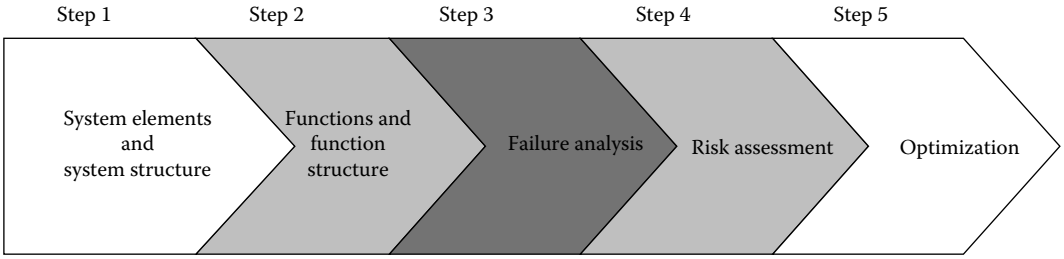


FIGURE 10.8 Five steps of the system—FMEA.

Step 3: Failure analysis. Determination of product defects—Failure, failure cause, failure effect, e.g., by means of coordination, previous complaints, previous FMEA.

Step 4: Risk assessment. Definition of avoidance and detection measures and evaluation of the defect severity (S) of the defect's effect, evaluation of the probability of occurrence (O) and of the probability of detection (D).

Step 5: Optimization. Derive the need for action, introduction of new, improved measures including target date and responsibility, controlling measures.

10.3 RELIABILITY ASSESSMENT IN EARLY DEVELOPMENT PHASES

Today's products are characterized by their ever-increasing complexity. As an example, consider how a gearbox has evolved over the years. It is noticeable that an increasing amount of software and electronic components has been integrated into the manual, the automatic, and the double clutch gearboxes. Since, on purchasing an automobile, the primary consideration of the customer is its reliability and not its appearance or price [1], this priority of ranking should be reflected by the manufacturer.

However, if one considers the actual reliability activities for mechatronic systems, the following shortfalls can be identified:

- Isolation of domains
- Consideration of reliability only in the latter phases of development; this entails great expense in modification and loss of time
- On the whole, only the reliability of components and assemblies is established
- The necessary data for the reliability calculations have, until now, been rarely available or deficient in quality and quantity

Consequently, there is a need for action that must focus on a mutual and holistic reliability system for mechatronic systems in early development phases.

10.3.1 Development Process and the V-Model

Generally, a procedure for a reliability analysis of mechatronic systems must be incorporated into the mechatronic development process. Only then, it is guaranteed that the reliability activities can accompany the development process.

The sequence of a mechatronic development process can be described with the help of the V-model from [12]. Since, in the framework of this process, the early development phases are decisive, the V-model's phase of system design is important (Figure 10.9).

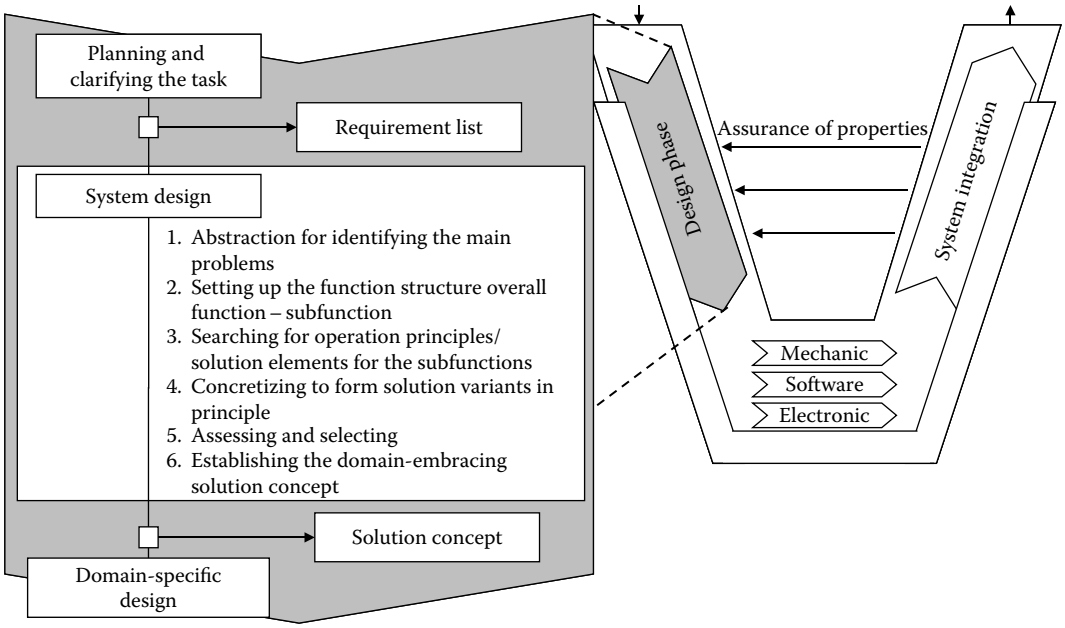


FIGURE 10.9 Extract of the V-model according to [12]. Activities in the system design.

10.3.2 Reliability Methodology for Mechatronic Systems in Early Development Phases

Starting from Figure 10.9, one has to ascertain how far the reliability can be incorporated into the system design. This approach is depicted in Figure 10.10.

To clarify the method, a component of an automatic transmission, the converter lockup clutch, is taken as an example (Figure 10.11). The approach is clarified by means of this example in the following sections.

10.3.2.1 Identification of Product Function/Malfunction

In the method's first stage, the product's top functions and its top malfunctions are specified similar to the FMEA. However, nonfunctional considerations like ease of repair or guarantee are not included. The top malfunctions are subsequently assigned to interception classes according to [16]. A reliability target value is allotted to each interception class. This value mainly results from consideration of marketing strategies.

For the example of the converter lockup clutch, the top functions are as follows:

- Close the clutch
- Open the clutch
- Transfer of torsional vibrations to the gearbox is prevented

The corresponding top malfunctions are

- Clutch does not close
- Clutch does not open
- Deletion of torsional vibrations is not guaranteed

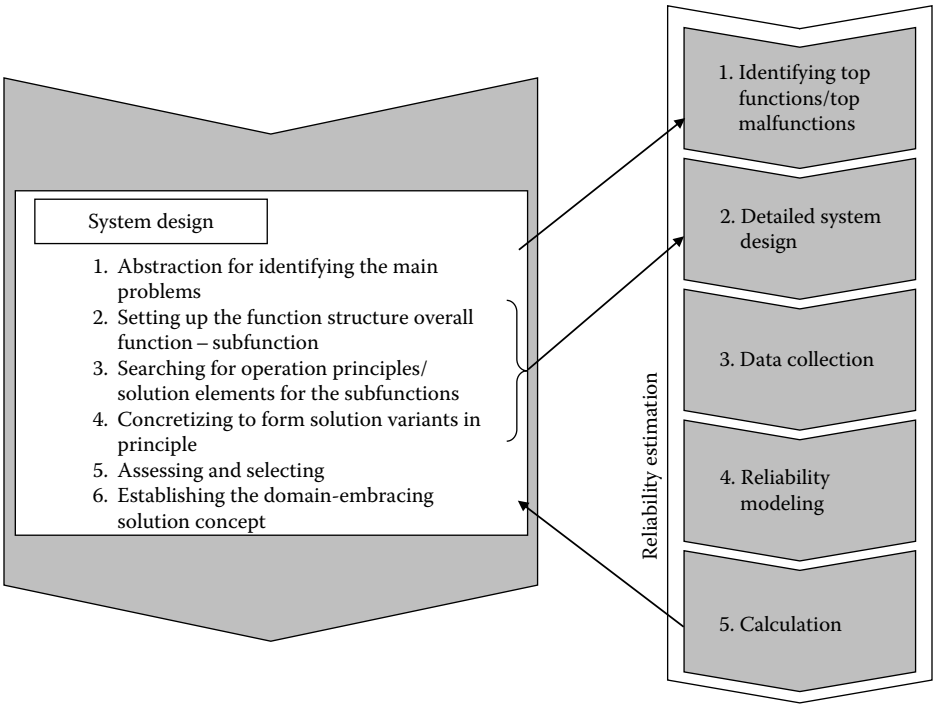


FIGURE 10.10 Approach for the reliability estimation in early phases of development which is incorporated to the first phase of the V-model—the system design. (From Jäger, P. et al., Zuverlässigkeitsbewertung mechatronischer Systeme in frühen Entwicklungsphasen. 4. Paderborner Workshop Entwurf mechatronischer Systeme, Universität Paderborn, Paderborn, Germany, 2006; Gäng, J. et al., *Determining Mechatronic System Reliability Using Quantitative and Qualitative Methods, ESREL 2007*, Stavanger, Norway, June 25–27, 2007.)

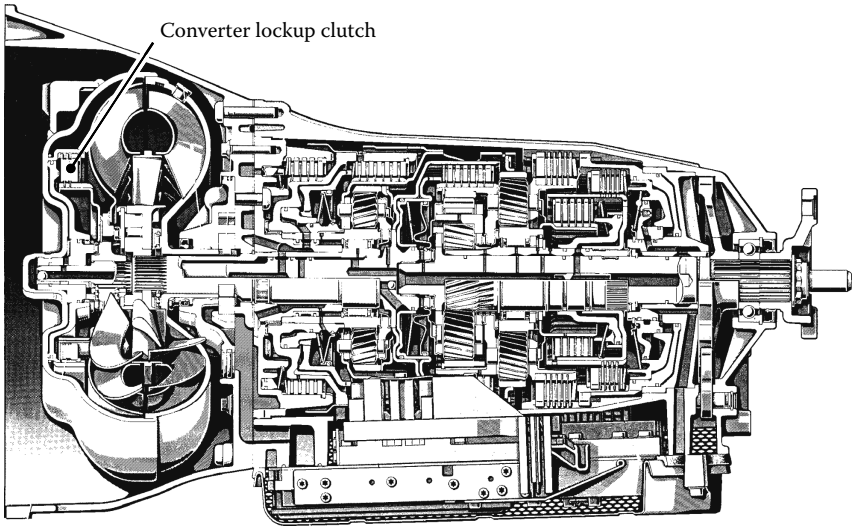


FIGURE 10.11 Automatic transmission with converter lockup clutch. (From Lechner, G. and Naunheimer, H., *Automotive Transmissions—Fundamentals, Selection, Design and Application*, Springer, Berlin, Germany, 1999.)

FIGURE 10.12 (a) Functional description of the system by means of use cases and (b) a first physical model generated on the basis of the use cases.

10.3.2.2 Detailed System Design

In the early development phases, information about the system already exists; this originates from previous or similar products. Using this knowledge, it is possible to annotate functional interrelationships using so-called use cases. Helped by the description, potential components for the converter lockup clutch can be singled out. These components can be adopted into a physical model that contains actuators and sensors necessary for the overall system (Figure 10.12). At the initial stage, it is sufficient that the electronics and the software are combined in information processing. However, they are to be differentiated at a later stage with the aid of the Unified Modeling Language (UML).

10.3.2.3 Data Collection

Reliability data for mechanical components is available from fatigue life models or simulations. The problems here are the uncertainties, such as manufacturing quality or effective engine performance. It is also possible to acquire data on electronic components using reliability models and standard elements with very narrow loading profiles. Problems are, for example, plug and socket connections. It is more difficult with software since there is, at present, no useful reliability model for software [17]. However, manufacturers often have intuitive and experience-based know-how. This expert knowledge should be exploited for the reliability model and data acquisition. In considering the software, the expert is asked about the input parameters which, using a Bayesian belief network, yields a defect density [13,17]. Despite the inaccuracy of the input data, it is possible for the experts to evaluate individual systems relative to each other. Here, the aim is no longer to determine a reliability parameter such as a B_{10} fatigue life (at this time, 10% of the system components are failed), but to ascertain the scope of development to attain the required reliability target. This scope of development is specified by the parameter α.

10.3.2.4 Reliability Modeling

Since the component configuration as well as the components themselves has a large influence on the overall reliability, it should be considered at this stage, which options are the best for the present system. When all the components have been pre-ascertained, these options can, using Boolean algebra, be structured in their respective primary defective functions (Figure 10.13) and, at a later stage, assigned to different complaint classes.

10.3.2.5 Calculation

By using a mathematical model of the system or, as an alternative, the Monte Carl method, with which it is possible to deal with inaccuracies, and the reliability target values defined in Section 10.1, it can be shown which of the investigated systems is comparatively the best (Figure 10.14).

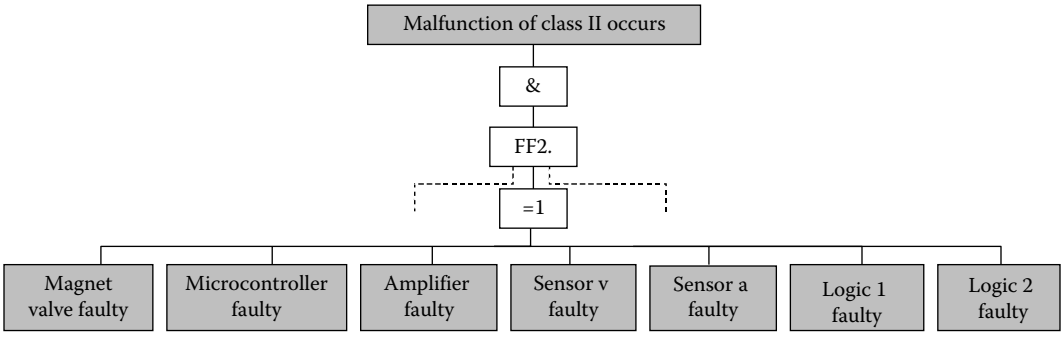


FIGURE 10.13 Fault tree.

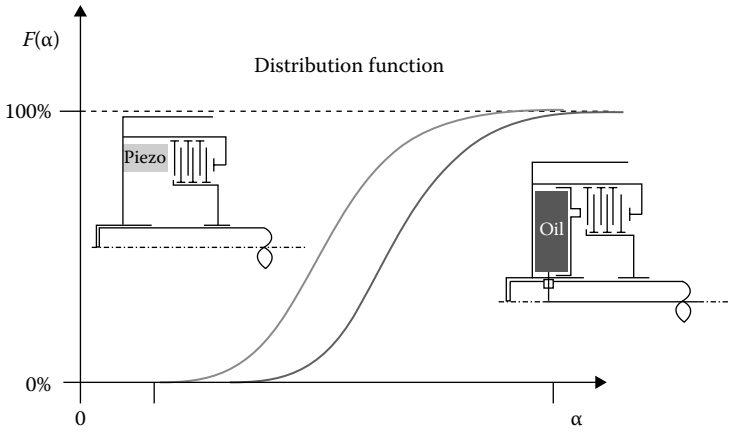


FIGURE 10.14 Comparison of different alternative solutions via quantification of the concept advantage.

10.4 NUMERICAL SENSITIVITY ANALYSIS AND RELIABILITY INVESTIGATION OF AN ACTIVE NOISE REDUCTION SYSTEM

Adaptronic systems react intelligently to variations of load conditions or their operating environment. Their functionalities include the active suppression of undesired noise and vibration, which are very promising for a number of automotive applications. Reliability of these complex adaptronic systems is hard to estimate for several reasons. Structural integration of sensor, actuators, and even control electronics, which are typical for adaptronic systems, requires a comprehensive analysis of all the physical domains involved to assess the complex system performance. Well-established methods for the design of mechatronic systems as outlined in the guideline VDI 2206 [12] focus on systems comprising of discrete electronic and mechanical components rather than integrated adaptronic systems which utilize multifunctional materials. A priori undetermined interactions between various influencing parameters, unknown fatigue properties of the multifunctional materials employed in sensors and actuators, and very complex system performance requirements make it difficult to predict the circumstances under which these adaptronic systems may fail. A sensitivity analysis (SA) of the comprehensive adaptronic system model has proven to be a valuable tool for the identification and assessment of scenarios that are relevant for system reliability [18,21]. By using an appropriately well-defined model, output and failure modes—that are designed as uncertain distributed input parameters, SA can be used to both identify and classify the importance of the different input values and to assess the reliability of the system with respect to each implemented failure mode.

As an example, the method is outlined for an active oil pan which is equipped with piezoelectric sensors and actuators to suppress structural vibrations. These are subject to more or less severe degradation and other failure modes which depend on operating conditions, applied control algorithm, and load-time history [19]. The performance of the active system, characterized by a defined reduction of noise emission in a particular frequency range, is simulated in different failure and fatigue scenarios. These include sensor and actuator degradation, debonding or position tolerances (misplacement), and malfunction. The influences of the design parameters and the different control strategies with respect to the performance robustness of the active oil pan system were able to be quantitatively analyzed through experimental as well as numerical sensitivity analyses [22,23].

10.4.1 Performance Evaluation

Figure 10.15 shows an FE model of the oil pan including piezoelectric (PZT) sensors and actuators (top) and the experimental setup (bottom) used for lab experiments. PZT patch actuators induce a bending deformation of the oil pan bottom which counteracts the vibration depending

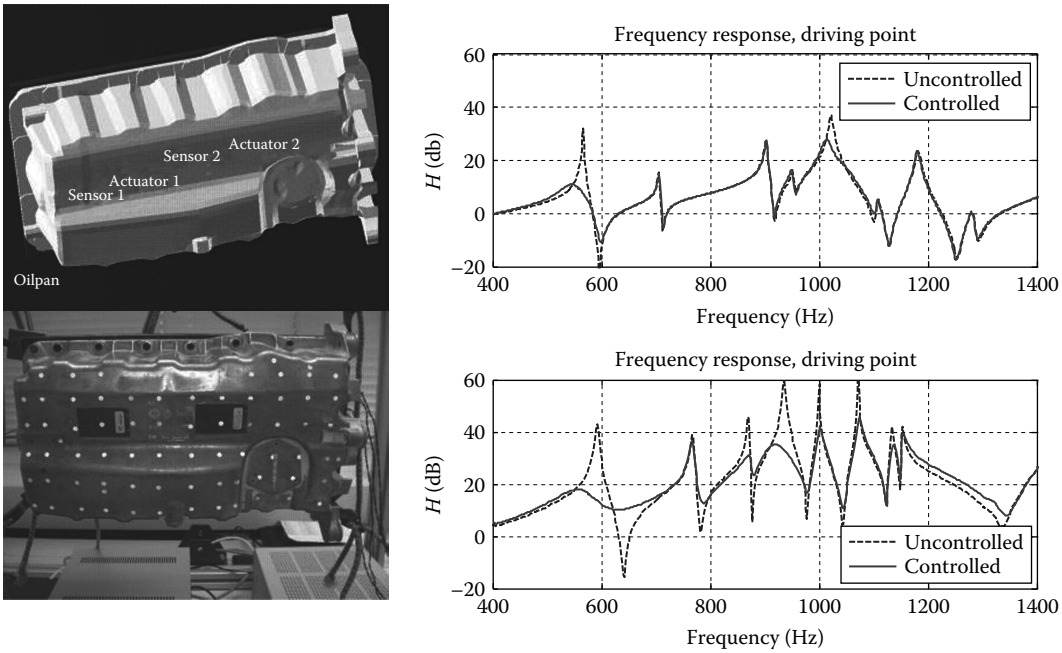


FIGURE 10.15 Numerical (top) and experimental (bottom) evaluation of the active system performance of oil-pan setup.

on the applied control algorithm. Frequency response plots of the oil pan subject to harmonic excitation are shown on the right-hand side of Figure 10.15 for numerical (top) and experimental (bottom) evaluation. The performance of the adaptive systems considered is typically defined as reduction of noise emission or vibration level of the uncontrolled (dashed line) vs. the controlled system (solid line) in some particular frequency range. Accordingly, the failure of the active system implies that the requirement for noise reduction within that range is not matched and the performance index falls below a specified critical limit. The performance index determines the difference in loudness or overall vibration level between the active (controlled) and passive (uncontrolled) system.

10.4.2 Sensitivity Analysis

A number of mathematical techniques are summarized under the expression SA [20]. When we mention SA in this context, we refer to variance-based sensitivity analysis. The aim of this type of SA is to determine the influence of uncertain, distributed input parameters on the output of a system. The parameters will generally be represented by distribution functions which are characterized by, for example, an expected value and a corresponding variance schematically depicted in Figure 10.16. As shown, distribution functions of the input parameters may be arbitrary, e.g., normal distributions, Weibull or uniform distributions, etc. and the variation of the input parameters results in a variation of the output function, e.g., the performance index. A simulation model serves to determine the output function of the system. In our case, the numerical FE model of the oil pan is used to generate the frequency response, which is then evaluated in designated frequency bands to generate a numerical value for the noise reduction performance of the system. The model could as well be an approximated response surface, identified from experimental data. Via the variance decomposition, the output function of the model is analyzed regarding the sensitivity of the output variance with respect to the input parameters.

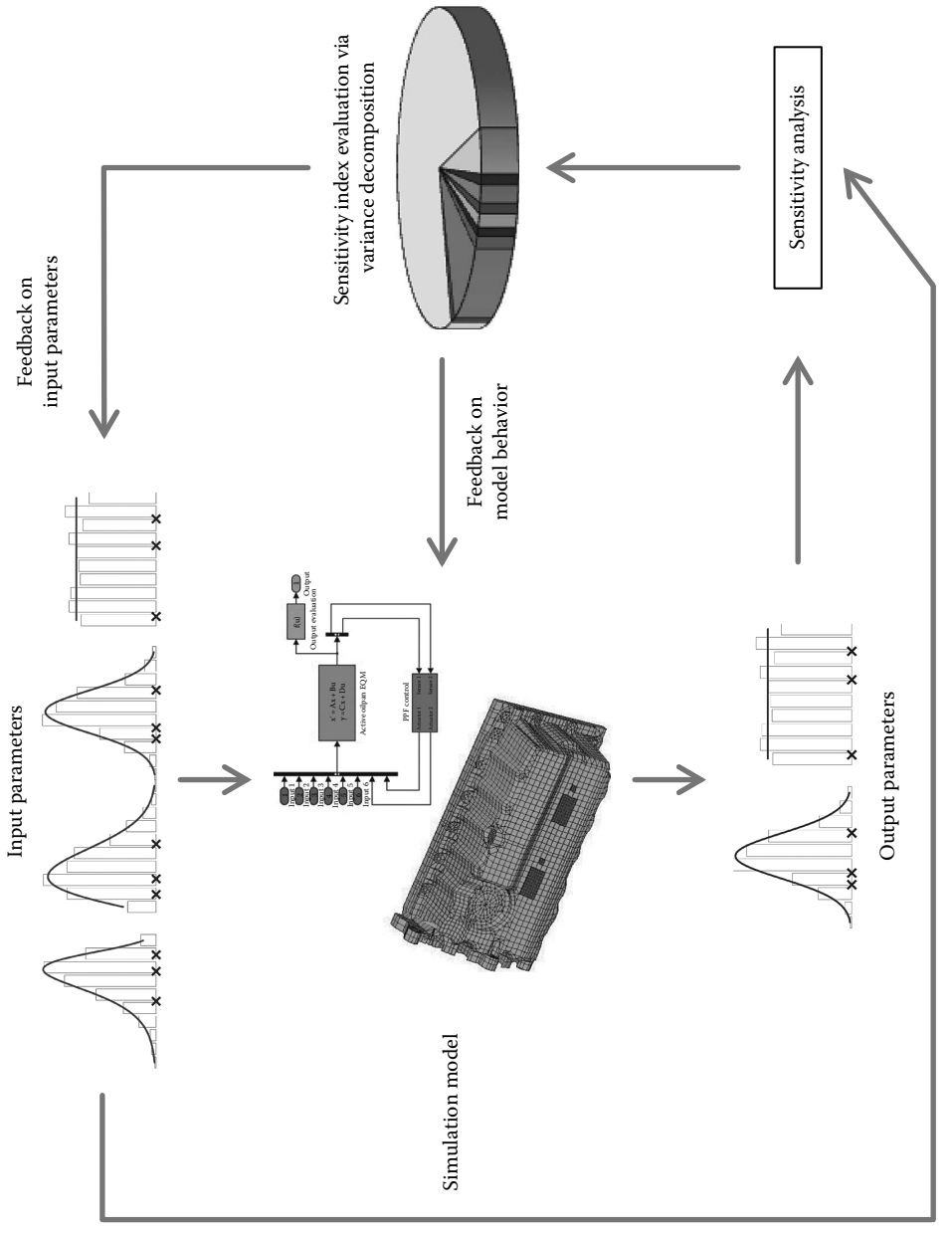


FIGURE 10.16 Variance-based sensitivity analysis of simulation model.

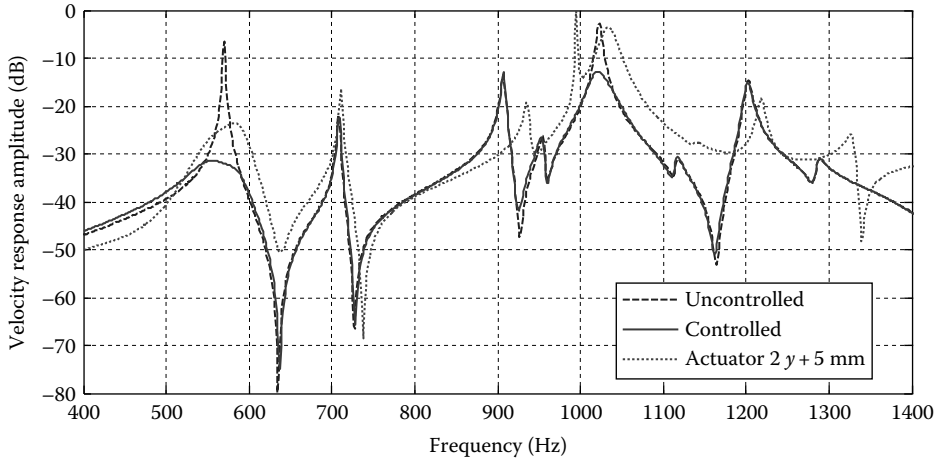


FIGURE 10.17 Vibration response of active system undergoing variation of actuator position.

Suitable sensitivity measures quantify not only the influence of single input parameters but also (higher order) interactions between them. Generally speaking, SA helps to find answers to the questions as follows:

- Which of the input parameters have the largest influence on variance of the output of the system?
- Which of the parameters have very little influence and may be eliminated from the model for simplification?
- How much variance of which input parameters is allowed to keep the output within specified limits?
- If there are "preferred" and "undesired" regions of the system output function, which of the input parameters have major influences on the output values in either region?

The absence of accurate data is a challenging issue when performing SA. For example, data describing degradation of actuators, accuracy of cutting tools, etc. are rarely available. Carefully designing and estimating the scattering of the input parameters are essential steps for the purpose of SA. This presents great potential for modern engineering design. Due to this large amount of uncertainty, the worst-case scenario is often taken into consideration, leaving enormous potential for an optimized reliable design.

As can be seen in Figure 10.17, the performance of the adaptive system dramatically changes with variations of the actuator position. The dashed line represents the calculated frequency response of the uncontrolled, the solid line the response of the controlled system, while the dotted line is the frequency response of the controlled system with one PZT actuator displaced from its nominal position by 5 mm.

The results of the SA carried out for actuator position and degradation are shown in Figure 10.18, according to the performance measures defined in Section 10.3.1. It is seen that a variation of the y-position of actuator 2 has the largest influence on the system performance.

10.4.3 Factors Influencing Reliability

Numerous quantities may be identified that influence the performance of the adaptive system and therefore may also contribute to the failure probability. For the presented method, the factors of interest are those that can be accounted for in the numerical model, such as

- Load spectra
- Design variations, tolerances

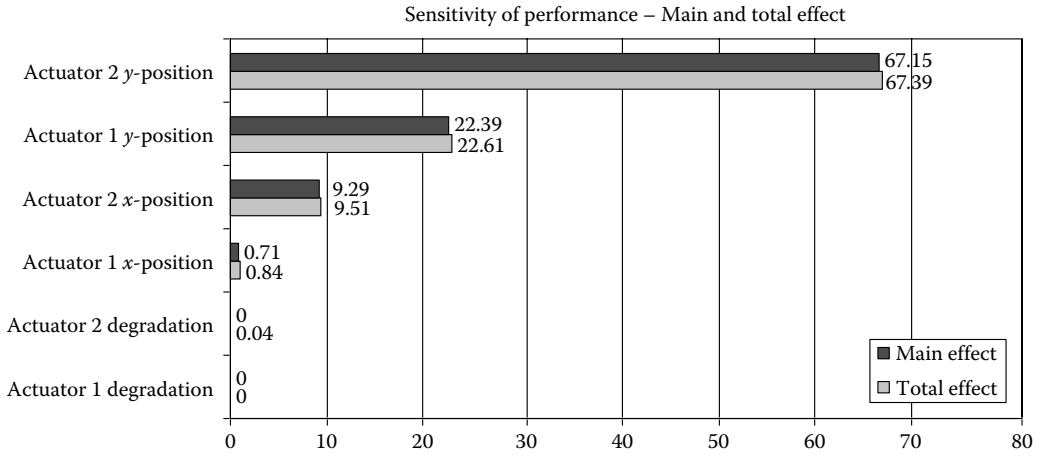


FIGURE 10.18 Results of SA for actuator position and degradation.

- Operating conditions
- Failure modes and degradation

All these quantities may be given as probability distributions and fed into the SA of the model. The outcome of the SA yields the importance of input quantities with respect to the variance of the output.

10.4.4 Lifetime Estimation

Based on the most relevant scenarios for system failure, which are identified via a SA, threshold values for input parameters can be given that assure system performance within specified limits. As the SA shows that variations in the position of actuator 2 have the largest influence on system performance, debonding or cracking of actuators should be one of the failure mechanisms to be closely investigated. In other system configurations, degradation of actuators and sensors has been identified to be likely cause of system failure. A focused reliability test aimed only at the most relevant failure mechanisms can now yield the required probability of a system component exceeding the input parameter threshold that depends on the applied number of load cycles. Based on this reliability function, a lifetime estimation of the system can be given. Collecting this reliability data is the subject of ongoing research.

10.5 CONCLUSION

Failure probability of components in mechatronic systems can be described using the Weibull distribution. Existing knowledge of component failure probabilities and the system structure may be combined to obtain a quantitative reliability prediction on a statistical basis. Systematic analysis of possible failure scenarios in an FMEA is a well-established procedure to identify and then reduce the vulnerability of a device with respect to various failure causes. With increasing complexity of mechatronic systems and shorter cycles of innovation, more reliability analysis must be based on simulation. Facing a multitude of model parameters influencing performance and reliability of mechatronic systems, current research efforts address the reduction of computational cost of these simulations.

REFERENCES

1. Bertsche, B. *Reliability in Automotive and Mechanical Engineering*. Springer, Berlin, Germany, 2008.
2. VDI 4001. *Reliability Terminology*. Beuth, Berlin, Germany, 2006.
3. Birolini, A. *Reliability Engineering—Theory and Practice*, 5th edn. Springer, Berlin, Germany, 2007.
4. Kececioglu, D. *Reliability and Life Testing Handbook*, Vol. 1. Prentice Hall, Englewood Cliffs, NJ, 1993.
5. Kapur, K.C. and Lamberson, L.R. *Reliability in Engineering Design*. John Wiley & Sons, New York, 1977.
6. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York, 1982.
7. Dhillon, B.S. *Mechanical Reliability: Theory, Models and Applications*. AIAA Education Series, Washington, DC, 1988.
8. VDA Band 4, Teil 2. *Sicherung der Qualität vor Serieneinsatz System*. FMEA, Frankfurt, Germany, 1996.
9. IEC 60812 ed.2. *Analysis techniques for system reliability—Procedure for failure mode and effects analysis* (*FMEA*). Geneva: International Electrotechnical Commission, 2006.
10. MIL-STD 1629A. *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*. Washington, DC: U.S. Department of Defense, 1980.
11. SAE J1739. *Potential Failure Mode and Effects Analysis in Design (Design FMEA) and Potential Failure Mode and Effects Analysis in Manufacturing and Assembly Processes (Process FMEA) and Effects Analysis for Machinery (Machinery FMEA)*. Society of Automotive Engineers, 2002.
12. VDI Guideline 2206. *Design Methodology for Mechatronic Systems*. Düsseldorf, Germany, 2004.
13. Jäger, P., Hitziger, T., and Bertsche, B. Zuverlässigkeitsbewertung mechatronischer Systeme in frühen Entwicklungsphasen. 4. Paderborner Workshop Entwurf mechatronischer Systeme, Universität Paderborn, Paderborn, Germany, 2006.
14. Gäng, J., Bertsche, B., Wedel, M., and Göhner, P. *Determining Mechatronic System Reliability Using Quantitative and Qualitative Methods. ESREL 2007*, Stavanger, Norway, June 25–27, 2007.
15. Lechner, G. and Naunheimer, H. *Automotive Transmissions—Fundamentals, Selection, Design and Application*. Springer, Berlin, Germany, 1999.
16. VDA Band 3, Teil 2. *Qualitätsmanagement in der Automobilindustrie—Zuverlässigkeitsabsicherung bei Automobilherstellern und Lieferanten, 3*. Auflage, Frankfurt, Germany, 2000.
17. Fenton, N.E. and Neil, M. A critique of software defect prediction models. *IEEE Transactions on Software Engineering*, 25(5), 1999, 675–689.
18. Wolf, K., Han, S., and Hanselka, H. Sensitivity based performance evaluation and reliability assessment of adaptive systems, *Proceedings of SPIE*, 6928, 2008, 69281A.
19. Nuffer, J., Schönecker, A., Brückner, B., Kohlrautz, D., Michelis, P., Adarraga, O., and Wolf, K. *Reliability Investigation of Piezoelectric Macro Fibre Composite (MFC) Actuators*. Adaptronic Congress, Göttingen, Germany, May 23–24, 2007.
20. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. *Global Sensitivity Analysis: The Primer*. Wiley, Chichester, U.K., 2008.
21. Li, Y., Pfeiffer, T., Nuffer, J., Bös, J., and Hanselka, H. *Experimental sensitivity analysis for robustness studies of a controlled system*, Journal of Smart Materials and Structures, 21 (2012) 064002.
22. Li, Y., Pfeiffer, T., Nuffer, J., Bös, J., and Hanselka, H. *Approaches to sensitivity analysis for system reliability study of smart structures for active vibration reduction*, Journal of Reliability, Quality and Safety Engineering, 19 (5) 1250025.
23. Li, Y., Pfeiffer, T., Nuffer, J., Bös, J., and Hanselka, H. *Experimentelle Sensitivitätsanalyse eines aktiven Struktursystems mittels D-Optimal Design*, DAGA 2012–38. Deutsche Jahrestagung für Akustik, Darmstadt, Germany, March 19–20, 2012.

Part II

Vehicle Concepts and Aerodynamics

11 Conceptual Design of Road Vehicles Related to Dynamics

Giampiero R.M. Mastinu

CONTENTS

11.1 Introduction 329
11.2 Brief History of Ground (Road) Vehicles 331
11.3 Basic Issues on Road Vehicle (Car) Design 335
11.3.1 Road Vehicle Dimensions 335
11.3.1.1 Wheelbase 336
11.3.1.2 Track 337
11.3.2 Suspension System 337
11.3.2.1 Lumped Mass Car Body 345
11.3.2.2 Ride Comfort and Road Holding Influenced by Main Vehicle Dimensions 346
11.3.3 Pneumatic Wheel 348
11.3.3.1 Load Carrying and Radial Stiffness 349
11.3.3.2 Rolling Resistance 353
11.3.4 Powertrain Concept 357
11.3.5 Traction and Braking 360
11.4 Active Safety 367
11.4.1 Handling 367
11.4.1.1 Tire Characteristics 368
11.4.1.2 Inertial and Geometric Data 371
11.4.1.3 Drive Axle and Torque Vectoring 376
11.4.1.4 Suspension Kinematics and Elastokinematics 380
11.4.2 Rollover 382
11.5 Vehicle Inertia Properties and Dimensions 383
11.6 Conclusion 388
References 390

11.1 INTRODUCTION

This chapter aims to present the theoretical foundations on which a vehicle is to be conceived and constructed, given the dynamic performance to be attained. The fundamental vehicle construction paradigms are presented and commented.

Apparently simple questions are addressed like why cars have four wheels at the four corners of the body or why suspension systems are basically composed by a spring nearby the ground (pneumatic tire) plus a spring and a damper between the wheel and the body. Other basic questions refer to the construction principles of pneumatic tires, active safety, and powertrain/driveline architecture.

In order to deal with such fundamental issues, theoretical investigations have been performed; some of the presented derivations are unreferenced. Only simple mechanical models, able to give general results, have been used. Alternatively, experimental data have been analyzed.

The theoretical analyses have shown that—necessarily—passive suspension systems must rely on a vertical massless spring in contact with the ground (pneumatic tire) to run properly on an irregular road. The hysteresis of such massless spring causes rolling resistance, such a phenomenon is explained in the chapter. Additionally, the compliance of the pneumatic tire in lateral direction, combined with the inertial and geometric characteristics of the vehicle, is the origin of many problems related to active safety (handling and cornering). Such problems are analyzed in this chapter from the conceptual design perspective.

The general conclusion of this chapter is that, given the constraints that have been set or inherited from the past and given the sustainable technologies available today, it is quite difficult to conceive road vehicles that are *completely* different morphologically from the ones we know, as today's road vehicles are the result of a huge optimization and system integration activity that started about 5000 years ago. The main result of the chapter is to have derived or gathered and emphasized a set of mathematical relations that underpin the design of today's road vehicles.

This chapter deals with a number of fundamental issues referring to the basic design of road vehicles, directly or indirectly related with dynamics. The aim of the chapter is to highlight why vehicles (and their subsystems) are constructed and shaped like we know them today.

A number of apparently naïve questions are addressed and answered with theoretical reasoning based on mathematical models. Such questions are as follows:

Is there a historical origin for today's dimensions of road vehicles?
Is today's architecture of road vehicles (cars) dictated by some specific reason?
Is there any necessary reason why the suspension system of road and off-road vehicles is combined by a pneumatic tire plus a spring and a damper?
Which must be the basic features of pneumatic wheels matching with their functions?
Which must be the basic features of road vehicle's suspension systems matching with active safety requirements?
Which must be the basic features of powertrain and driveline?
How do the geometric and inertial characteristics of a vehicle influence the active safety?
Which are the mass properties of road vehicles?

To answer the aforementioned addressed questions, simple models and analytical formulae have been used, whenever possible, to make the analyses as general as possible.

There are in the literature many books that deal with fundamental topics related with the performance of road or off-road vehicles and their subsystems [1–18,108]. In such books, the dynamic behavior of vehicles is explained in a very accurate and informative way by means of refined mathematical models that take into account the *existing* mechanical systems (vehicles and their subsystems). In this chapter, an attempt is made to explain why—if one or more reasons exists—the vehicles (and their subsystems) are constructed and shaped as the common experience would suggest. The mathematical models used in the literature to describe the motion of an existing vehicle are used here to investigate why such a vehicle had to be conceived and constructed in that particular way in order to attain some prescribed dynamic performance.

The addressed issues refer to the probable historical origin of today's dimensions and architecture of road vehicles, to the car body architecture as influenced by basic functional and technological requirements, to the arrangement of the suspension of road vehicles that is composed by a pneumatic tire plus a spring and a damper (or by an active system), to the basic dimensions and features of pneumatic wheels matching with their functions, to the driveline layout related to (urban) mobility and power losses, to the geometric and inertial features of cars and vans influencing active safety, to the basic features of road vehicle's suspension systems matching with active safety requirements, and to the mass properties of road vehicles.

11.2 BRIEF HISTORY OF GROUND (ROAD) VEHICLES

It is believed that ground vehicles became an important tool for carrying heavy goods when humans started agricultural activities and stopped from being nomadic hunters–gatherers [19]. This seems to be happened roughly in the time interval 15,000–5,000 BC, mainly in the Mesopotamic area. An important factor that contributed to the development and employment of ground vehicles was the possibility to exploit locomotion energy provided by domesticated animals, such as onagers (a kind of wild ass). This might explain why, before European colonization, in the Americas vehicles were not in use; actually animals strong enough were not available. The early vehicles in Asia and Europe were simply sleds or log sleds (Figure 11.1), running on grass or sand/rocky surfaces, sloped (mainly) or horizontal. The architecture of the sleds was strictly related to the power and the force available (for today's horse maximum values that are about 0.5 kW and 600 N, surely the early horses, being smaller [20], provided much lower mechanical power).

It is believed that, to reduce the friction and thus the traction force, rollers (e.g., logs) were used. Using a plank to carry a load would have suggested the idea of connecting the logs to the plank, thus leading to the invention of the wheel (Figure 11.2) [19,21]. The wheel has been one of the main inventions of the mankind (not a discovery). No structure is similar to a wheel in nature.

The history of wagons and carts has been documented in [1,19,21–28]. One of the very first vehicles employed for war is depicted in Figure 11.3, dating back about 2500 BC. Two-piece solid wheels can be seen.

The architecture of the early carts running in the Mesopotamic area (and possibly later in the Mediterranean area before the fifth century BC) refers to two-axle vehicles (wagons) with non-steering wheels. The observation that wheels dug ruts during repeated runs on the same path eventually suggested to build paved roads with ruts deliberately made. This happened in Ur, Babylon, and ancient cities of Assyria and implies that trains came before cars. The Prophet Isaiah in the eighth century BC refers about Assyria that the crooked were made straight and the rough places plain [22].

The importance of the road in defining the size (track) of carts was clearly known by Romans. Romans built many roads departing from Rome toward the borders of the empire, as the army—based in Rome—had to reach the far crisis areas as soon as possible. The roads were built to allow standard war carts to run without interfering with the other ones and produce traffic troubles. It is argued (no written documents stating this are today available) that the size

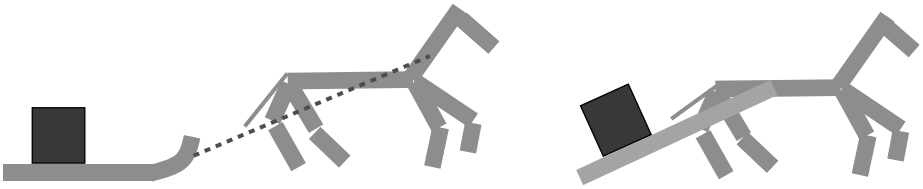


FIGURE 11.1 Schematic view of a sled (a) and a log sled (b) towed by an onager.

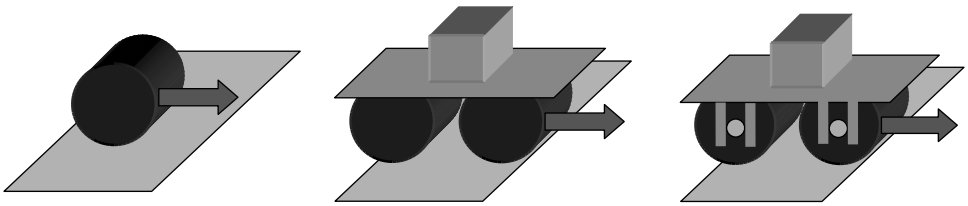


FIGURE 11.2 Logical process possibly leading to the invention of the wheel. Two-axle vehicles (wagons) were firstly developed. See Figure 11.6 showing how logs (could) have been linked to the plank.



FIGURE 11.3 The Standard of Ur, British Museum, 2600–2400 BC.

(cross width) of the cart was defined by the cross width of two horses pulling side by side the cart itself, as shown in Figures 11.4 and 11.5. The connection of the horses to the cart (harness system [24]) was made by means of proper structures resting on the necks. (There is some literature dealing with Roman harness system. In [19,21], authors state that until the Middle Age and after, the maximum power delivered by horses was about 0.12 kW, due to horses towing

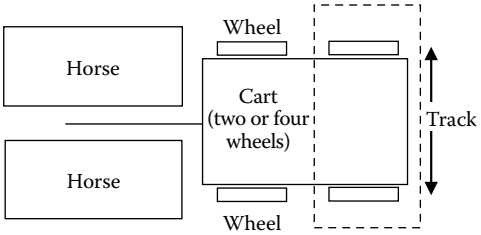


FIGURE 11.4 Hypothesis on Roman cart architecture. The cart track was related to the width of two horses. The cart had two or four wheels (wagon). Notice the only shaft along the middle axle of the cart.



FIGURE 11.5 Ruts left by ancient Roman carts in Pompeii, Italy, 59 BC. The track is about 1.4 m.

vehicles by a yoke resting on the throat and limiting the oxygen inspired by the animal. Such a thesis seems wrong according to [20].)

The track of early road vehicles (about 1.4 m) that were developed in Europe at the end of the nineteenth century AD (Benz, 1885, 1889) is related to the track of carts running in Europe since the Renaissance (1400 AD), when the ancient Roman roads (width about 3 m) were reused for trade. So we may argue that today's cross width of cars (track) is somehow related to the layout of very ancient carts running in Europe 2000 years ago.

The first known steering wagons were the Celtic carts found in Dejbjerg, Denmark, and dated first century BC [19]. Such carts were possibly used for ceremonies and they were obtained simply by connecting two single-axle carts. There is no evidence in carvings that ancient Roman four-wheel carts had a steering front axle [19]; however, after relatively recent studies, Röring [27] concluded that Roman wagons did have a pivoting front axle. Surprisingly, in the Middle Age in Europe, wagons appear not having a pivoting front axle. Pivoting front axles were reintroduced in Europe after 1300 AD and then extensively used [19,21]. Wagons with a pivoting front axle are easily recognized as they often have the front wheels smaller than rear ones.

Suspended vehicles were widely used in the Roman Empire [27], as an extreme luxury. Imperator Adrianus had a suspended system for playing with dice on his wagon [29]. Again in the Middle Age, the wagons were not suspended anymore. A suspension system very similar to the Roman one, based on leather strips, was reinvented during the mid-1400 AD in Hungary [19], apparently in the town of Kocs, which gave the name to suspended vehicles named "coaches."

During the Roman age, weight limits were specified for the Roman Imperial Post service (*De Cursu Publico*, Codex Theodosianus) and were related to the different types of vehicles (carts) [20,29]:

Angaria—1500 Roman lb, 492 kg
Raeda (wagon—four wheels, passenger)—1000 Roman lb, 330 kg
Carrus (wagon—four wheels, goods)—600 Roman lb, 198 kg
Vereda—300 Roman lb, 99 kg
Birota (cart—two wheels)—200 Roman lb, 66 kg

For more information on different types of carts in the Roman age, see [29]. The weights were limited also to avoid damaging the roads.

In ancient times, the shape of the wheels was strictly related to the stress they were subjected and to the unevenness they had to traverse. Taking into account the Hertz contact theory, a guess can be made on how the shape of wheels had evolved (actually Hertz developed his theory in the late nineteenth century eventually referring to railway wheels). Let us consider [30] Young's modulus of wood $E = 10$ GPa, Poisson's modulus of wood $\nu = 0.3$, the admissible compressive stress of wood $\sigma_{adm} = 8$ MPa (for loading perpendicular to grain), and a load $F_z = 1000$ N. The Hertz maximum stress P_0 in the contact area of a transverse log like the one depicted in Figure 11.6 (radius $R = 0.1$ m, length $l = 1$ m) is

$$P_0 = \left(\frac{F_z E^*}{l \pi R} \right)^{1/2} = 6\,\text{MPa}$$

where

$$\frac{1}{E^*} = \frac{1 - \nu^2}{E}$$

The result of the computation shows that such a structure (transverse log) may safely work as $P_0 < \sigma_{adm}$. Small-radius wheels are unable to run on obstacles, so bigger-diameter wheels had to be, and

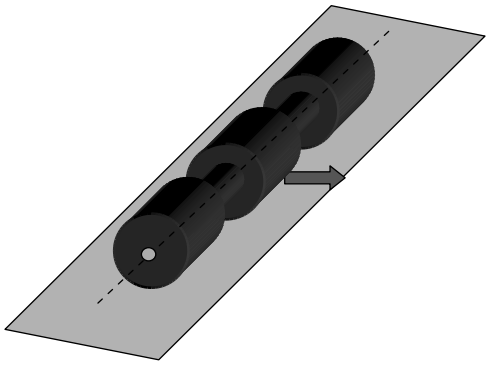


FIGURE 11.6 Log with grooves for linking it to an above plank on which goods could be carried.

were, adopted. Such an increase of diameter, due to weight reasons, implied *thin wheels*; fortunately, the Hertz contact pressure decreases as the radius increases. A 0.5 m diameter wheel (R = 0.25 m, width l = 0.18 m) implies the Hertz contact pressure to be about 8 MPa. Wheels dating 2750 BC had a diameter of approximately 0.5 m [19]. Diameters ranging from 0.5 to 1 m were in use in ancient times [19], up to the end of the nineteenth century. To decrease the weight of the wheels, spoked structures were used by Egyptians (1500 BC) [19] and much later by the Celts (200–100 BC). The manufacture of spoked structures requires an in-depth knowledge of technology and material science; a comprehensive explanation on how wheels were manufactured can be found in [19,20].

To limit the wear of the wheels, a number of treads were attached to the wheel rims [19] (see, e.g., the carvings in the Louvre museum, Near Eastern Antiquities pavilion). Celts, Greeks, and Romans mounted metal rims on wheels. Romans used steel rims (the steel was similar to today's low carbon steel) in only one piece hot shrink fitted on the wooden wheel (such a technology was rediscovered after the Middle Age) [19,20]. Noise produced by steel wheels was a major concern in the ancient city of Rome [29].

The wheels were at first connected to each other rigidly (like today's railway wheelsets). Greeks had such a construction scheme [19], but the mentioned earlier Egyptian carts had independently rotating wheels. Celts used cylindrical roller bearings (first century BC [19]), but up to the twentieth century, all of the wheel bearings were plain (roller or ball bearings were introduced at the beginning of the twentieth century).

Spoked wheels were prone to failures due to shocks acting in lateral direction. In order to diminish the flexure in the spokes due to such shocks, starting from the fifteenth century, a number of engineering solutions were adopted. Such solutions are depicted in Figure 11.7. One solution was to

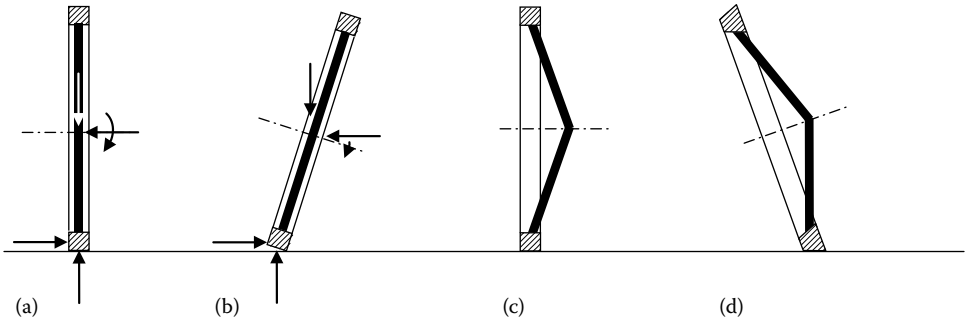


FIGURE 11.7 Cross section of spoked wheels, (a) spokes subject to flexure caused by a lateral shock, (b) spokes less subject to flexure (with respect to case (a)) due to the cambered wheel, (c) cambered spokes (less subject to flexure due to shock but subject to flexure due to radial load), and (d) sloped wheel axis (less subject to flexure due to shock and less subject to flexure due to radial load).

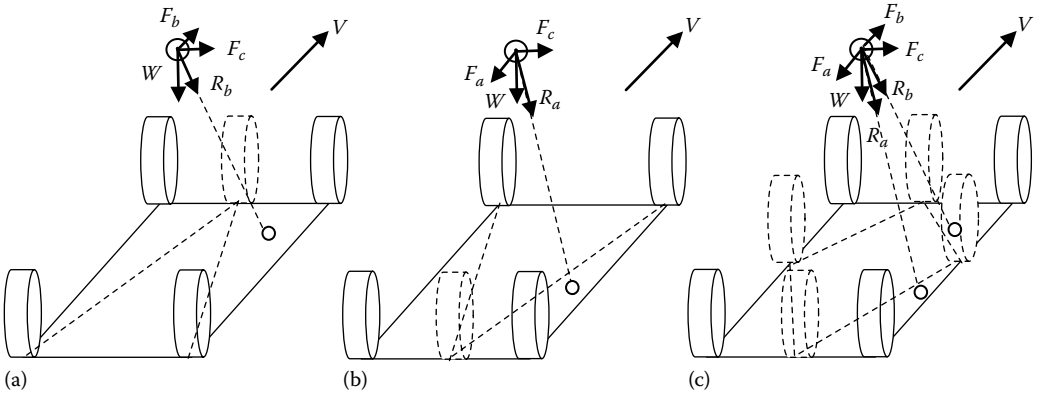


FIGURE 11.8 Comparison of rollover propensity of non-tilting vehicles with different arrangements of the wheels. The arrangement of four wheels at the four corners of the base rectangle is always better with respect to other arrangements with three wheels or with four wheels. (a) Three-wheeled vehicle during braking and cornering, (b) three-wheeled vehicle during acceleration and cornering, and (c) four-wheeled vehicle (rhomb arrangement) during braking or acceleration while cornering. Symbols: W weight, F_a acceleration force, F_b braking force, F_c centrifugal force, R_a resultant of W, F_a, F_c forces, and R_b resultant of W, F_b, F_c forces.

set the wheel axis not parallel with respect to the ground [21] (Figure 11.7b), but this implied a very large track. To reduce the track, the spokes were sloped (cambered) with respect to the horizontal (Figure 11.7c), but this implied the flexure in the spokes due to the payload. A sloped wheel axis together with a wheel with sloped (cambered) spokes was a solution (Figure 11.7d). Later carts or wagons had big spoked wheels (about 1.5 m diameter) to improve the running on road irregularities.

Historically, vehicles were conceived since the very beginning (see Figure 11.3) with four wheels at the corner of a flat rectangular plank used as vehicle body. This appears to be still today the more rational way to define the position of the wheels; actually for static reasons, this is the better way to prevent rollover (non-tilting vehicles are considered). Figure 11.8 shows the advantages of such a layout considering forces applied to the center of gravity during braking or acceleration while cornering. If the resultant force at the center of gravity points outside the area defined by the lines connecting the wheel–ground contact points (respective central points of tire–ground contact areas), the vehicle becomes prone to roll over. At least one attempt to place wheels at the vertexes of a rhomb has been made [31]. A vehicle with rhomboid arrangement of the four wheels suffers similar rollover problems of three-wheeled vehicles, unless tilting is implemented [32].

11.3 BASIC ISSUES ON ROAD VEHICLE (CAR) DESIGN

11.3.1 Road Vehicle Dimensions

The track of early automobiles was possibly inherited from ancient Roman standards (see Section 11.2 and also [112]). Nowadays, there are national and international laws, as well as standards that rule not only the track but also the length and the height (as well as the mass) of vehicles allowed to run freely on public roads (see ISO 3833 [33]).

In [3,4,34,35], the basic issues referring to the conceptual definition of the dimensions of road vehicles—with particular reference to automobiles—are given. According to [35], the parameters for the design of a vehicle (automobile) can be classified into the three groups, *anthropometry, customer requirements*, and *technical equipment*.

The anthropometric data useful for the design of automobiles refer primarily to the ones of the seated person (references for the human dimensions are given in [4,35,36]); the automobile is actually designed (somehow "tailored") around the passengers. The starting point is defining the

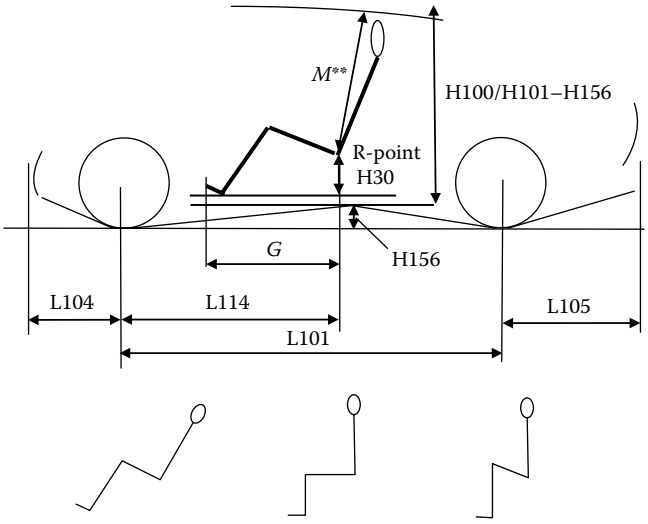


FIGURE 11.9 Main dimensions of an automobile (referring to SAE J1100 and/or [35]) and different body postures influencing directly the body shape of a car.

height of the so-called H point (hip point) located at the rearmost position of the seat (R point) [36–38] (Figure 11.9) with respect to the floor (H30). A rough relationship exists between H30 and G (Figure 11.9; H30 = 1770 – 1.622 × *G*, [35]). Higher H30 implies smaller G (H30 ranges between 170 and 390 mm [35]). Experience and research activity [37] show that the ride comfort (see Chapter 36; [38–41]) is related to the posture, so the definition of *G* and H30 is of crucial importance for vehicle character. The angles between body segments that are expected to provide posture comfort are given in [5,6,42–45] mainly referring to static condition. The height *M*** is set (according to [35]) for some vehicles of a given manufacturer equal to 993 mm (Figure 11.9). In [35], details are given on pedal positioning, steering wheel positioning, and cross width dimensions.

As the passenger room is defined, the internal boot dimensions have to be defined (according to customer requirement); if possible, as a multiple of the cuboid defined by ISO 3832, this helps preventing a too irregular inner shape of the boot. In front engine road vehicles (the greatest majority of cars), the front length (from front passenger feet to the foremost point) is defined by the bigger motor type that has to be fitted into the engine compartment [35]. The total length of the vehicle is thus defined according to such requirements.

Additional details on the interiors, the doors, and the visibility of an automobile can be found in [3,5,6,46–51]. Additional details on how designing the shape of an automobile can be found in [3,5,6,52–60].

From the analysis of the literature, it appears that, starting from the requirements referring to anthropometry, customer, and technical equipment, *there is not a well-defined constraint to set the wheelbase and the positioning of the front and rear axles* (L101, L104, and L105 in Figure 11.9). So these parameters are chosen according to the designer's preferences. There are many different possible preferences; the considerations based on vehicle dynamics are addressed as follows.

11.3.1.1 Wheelbase

The wheelbase length has effects on a number of vehicle performances, namely, acceleration and braking (see Section 11.3.5), handling (see Sections 11.4.1.2 and 11.4.1.3), and ride comfort (see Section 11.3.2.2). Also, (off-road) mobility performance is greatly influenced by the wheelbase (see Chapter 12). Additionally, the setting of the wheelbase has direct effects on the mass and stiffness of the car body (see Chapter 26) and on inertia properties (see Section 11.5). There is not a generally accepted rule to set the wheelbase of an automobile, due to the many different dynamic performances it affects.

11.3.1.2 Track

Also the track, as the wheelbase, has effects on a number vehicle performances, namely, rollover (see Section 11.4.2), handling (see Sections 11.4.12 and 11.4.1.3), and ride comfort (see Section 11.3.2.2).

11.3.2 Suspension System

Let us consider how the wheels can be connected to the car body and whether or not a suspension system has to be fitted into the vehicle. Experience obviously states that a suspension system is necessary; here, we will focus on the basic paradigm of suspension systems. Let us consider a vehicle like an ancient single-axle cart, for example, the one represented in Figure 11.10a. Such a vehicle does not have any suspension, that is, it is unsprung. As the road profile is rough, vibrations are excited, mainly in the vertical direction. The question arises if, given a certain level or road irregularity, there is a limit speed, above which such a vehicle cannot run, due to the wheel leaving the ground.

The statistical description of random road profiles has been given in [2,61–63]. Referring to Figure 11.10, the displacement ξ (road irregularity) may be represented by a random variable defined by a stationary and ergodic stochastic process with zero mean value and Gaussian distribution [61,64]. A concise but complete summary of the theory of random vibration and related stochastic processes is given in [65]. The mean square spectral density of an uneven road profile is given by

$$S_{\xi 1} = \frac{A_b v}{\omega^2} \ [\mathrm{m^2\,s}] \tag{11.1}$$

or

$$S_{\xi 2} = \frac{A_v s_c}{s_c^2 + \omega^2} \ [\mathrm{m^2\,s}] \tag{11.2}$$

where

v is the vehicle speed

ω is the circular frequency (rad/s)

A_b, A_v, s_c are the parameters referring to the irregularity (typically for medium roads) $A_b = 1.4 \times 10^{-5}$ m, $a = s_c/v = 0.4$ rad/m, and $A_v = 3.5 \times 10^{-5}$ m^2; a comprehensive treatment is reported in [2,66]

Equation 11.1 represents the simplest way to describe with sufficient accuracy a random road profile, but it has the drawback that the irregularity amplitude grows to infinity as ω approaches zero. Such a drawback is absent in Equation 11.2. For reasoning about basic issues, Equation 11.1 is often sufficient.

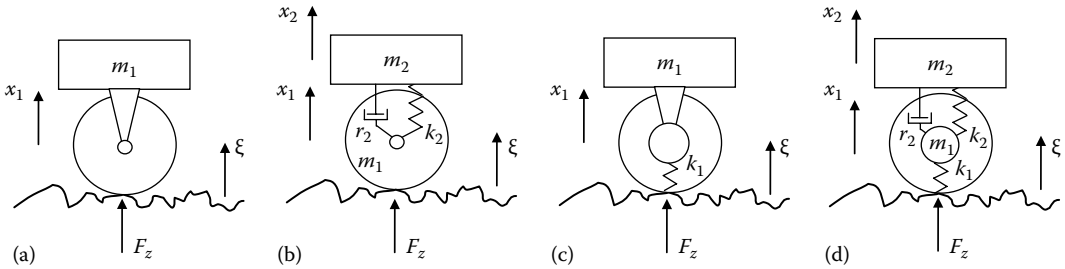


FIGURE 11.10 System models representing vehicles running on an uneven road profile. (a) 1 d.o.f. unsprung vehicle model, (b) 1 d.o.f. sprung vehicle model with stiff wheel, (c) 1 d.o.f. unsprung vehicle model with elastic wheel, and (d) 2 d.o.f. sprung vehicle model with elastic wheel.

The performances that are relevant for studying the vertical vibrations of the unsprung vehicle in Figure 11.10a are

- The body vertical acceleration (represented by $\ddot{x}_1$ in Figure 11.10a) and related to passenger discomfort and rattling of carried goods
- The force at the wheel–road contact area (represented by F_z)

Notice that for the vehicle model in Figure 11.10a, $m_1\ddot{x}_1 = F_z$, that is, the body vertical acceleration and the vertical force at the ground are strictly related. As the distribution of the irregularity ξ is Gaussian, the distribution of $\ddot{x}_1$ and F_z is Gaussian (if the wheel does not leave the ground). Referring to Tchebycheff's inequality, for a Gaussian distribution, if $3\sigma_{F_z} = F_{z,static} = m_1 g$ (g gravity acceleration), there is the probability that 0.3% of the time the wheel leaves the ground, that is, $F_z < 0$. This condition of the wheel not touching the ground is obviously a condition to be avoided. In order to compute the standard deviation of the vertical force σ_{F_z} (and the standard deviation of the vertical acceleration $\sigma_{\ddot{x}_1}$), let us consider that $\ddot{\xi} = \ddot{x}_1$ and that

$$S_{\ddot{\xi}1} = \omega^4 S_{\xi 1} = \omega^4 \frac{A_b v}{\omega^2} = \omega^2 A_b v \tag{11.3}$$

so (see [2,65,67])

$$\sigma_{F_z} = m_1 \sigma_{\ddot{x}_1} = m_1 \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^4 S_{\xi 1} d\omega} = m_1 \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^4 \frac{A_b v}{\omega^2}\, d\omega} = m_1 \sqrt{\frac{1}{2\pi} \left[A_b v \frac{\omega^3}{3} \right]_{-\infty}^{\infty}} = \infty \tag{11.4}$$

Additionally, for

$$S_{\ddot{\xi}2} = \omega^4 S_{\xi 2} = \omega^4 \frac{A_v s_c}{s_c^2 + \omega^2}$$

$$\sigma_{F_z} = m_1 \sigma_{\ddot{x}_1} = m_1 \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^4 S_{\xi 2} d\omega} = \infty \tag{11.5}$$

For an unsprung vehicle represented by the model in Figure 11.10a, the standard deviations of the vertical force σ_{F_z} and the standard deviation of the vertical acceleration $\sigma_{\ddot{x}_1}$ are, respectively, *unbounded*. This implies that, unless the vehicle speed is vanishing or the road profile is perfectly smooth (i.e., $A_b \to 0$ or $A_v \to 0$), both high vertical forces (and wheels missing the contact with the ground) and high vertical accelerations can arise. So, theoretically, for an unsprung vehicle, the limit speed—to avoid the wheel leaving the ground—is *zero*. Actually, it is well known that unsprung vehicles can run only at very low speed.

Let us now consider the vehicle system model depicted in Figure 11.10b, representing a sprung vehicle, that is, a vehicle fitted with a suspension system. The equation of motion of such a simple 1 d.o.f. system is

$$m_2\ddot{x}_2 + r_2\dot{x}_2 + k_2 x_2 = r_2\dot{\xi} + k_2\xi \tag{11.6}$$

the frequency response of the amplitude $x_2 = X_2 e^{j\omega t}$ due to the inputs $\xi = \Xi e^{j\omega t}$, $\dot{\xi} = i\omega t \Xi e^{j\omega t}$ is, for $\Xi = 1$

$$X_2 = \frac{j\omega r_2 + k_2}{-\omega^2 m_2 + j\omega r_2 + k_2} = H_{x_2} \tag{11.7}$$

and

$$\ddot{X}_2 = -\omega^2 \frac{j\omega r_2 + k_2}{-\omega^2 m_2 + j\omega r_2 + k_2} = H_{\ddot{x}_2} \tag{11.8}$$

For $S_{\xi 1}$ excitation, the standard deviation of $\ddot{x}_2$ is [2,67]

$$\sigma_{\ddot{x}_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \left|H_{\ddot{x}_2}\right|^2 S_{\xi 1} d\omega} =$$

$$= \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} (j\omega)^2 \frac{j\omega r_2 + k_2}{(j\omega)^2 m_2 + j\omega r_2 + k_2} (-j\omega)^2 \frac{(-j\omega) r_2 + k_2}{(-j\omega)^2 m_2 + (-j\omega) r_2 + k_2} A_b v \frac{1}{j\omega} \frac{1}{(-j\omega)} d\omega} = \infty \tag{11.9}$$

In fact, rearranging Equation 11.9

$$\sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} A_b v \left[j\omega \frac{j\omega r_2 + k_2}{(j\omega)^2 m_2 + j\omega r_2 + k_2} \right] \left[(-j\omega) \frac{(-j\omega) r_2 + k_2}{(-j\omega)^2 m_2 + (-j\omega) r_2 + k_2} \right] d\omega} =$$

$$= \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} A_b v A(\omega) A(-j\omega) d\omega}$$

one can easily see that $\lim_{i\omega \to \infty} A(\omega) = -r_2/m_2$, that is, the function $A(\omega)$ to be integrated has a horizontal asymptote for $\omega \to \infty$ that implies $\sigma_{\ddot{x}_2} = \sqrt{(1/2\pi)\int_{-\infty}^{\infty} \left|H_{\ddot{x}_2}\right|^2 S_{\xi 1} d\omega} = \infty$.

With similar reasoning, for $S_{\xi 2}$ excitation, the standard deviation of $\ddot{x}_2$ is

$$\sigma_{\ddot{x}_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \left|H_{\ddot{x}_2}\right|^2 S_{\xi 2} d\omega} = \infty \tag{11.10}$$

The acceleration of mass m_2 is *unbounded*; this is due both to the power spectral density of the excitation ($S_{\xi 1}$ or $S_{\xi 2}$) that decreases according to $1/\omega^2$ and to the presence of the damping r_2.

Similarly, to what was written for the model in Figure 11.10a, for the model in Figure 11.10b, we can write $m_1 \ddot{x}_1 + m_2 \ddot{x}_2 = F_z$, so (from 11.3–11.9)

$$\sigma_{F_z} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \left(m_1^2 \omega^4 + m_2^2 \left[(j\omega)^2 \frac{j\omega r_2 + k_2}{(j\omega)^2 m_2 + j\omega r_2 + k_2} \right] \left[(-j\omega)^2 \frac{(-j\omega) r_2 + k_2}{(-j\omega)^2 m_2 + (-j\omega) r_2 + k_2} \right] \right) S_{\xi 1} d\omega = \infty}$$

The same result is found for $S_{\xi 2}$ excitation. The standard deviation of the vertical force σ_{F_z} is *unbounded*. The *sprung* model in Figure 11.10b suffers the same problem of the *unsprung* model in Figure 11.10a, that is, the limit speed—to avoid the wheel leaving the ground—is *zero*.

Surprisingly, for the sprung vehicle model in Figure 11.10b, the standard deviation of the vertical body acceleration $\sigma_{\ddot{x}_2}$ and the standard deviation of the vertical force at the ground σ_{F_z} are *unbounded*. On the contrary, the amplitude response σ_{x_2} is bounded:

$$\sigma_{x_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty} \left|H_{x_2}\right|^2 S_{\xi 1} d\omega} = \sqrt{\frac{1}{2}\frac{A_b\left(m_2 k_2 s_c r_2^2 + m_2\left(m_2 s_c + r_2\right)k_2^2\right)}{m_2 k_2\left(m_2 k_2 s_c + \left(m_2 s_c + r_2\right)k_2\right)}} \tag{11.11}$$

Let us now consider the vehicle model in Figure 11.10c. It behaves almost like the one in Figure 11.10b, provided that $x_1 = x_2$, $k_1 = k_2$, $r_2 = 0$.

For $S_{\xi 1}$ excitation, the standard deviation of $\ddot{x}_1$ is

$$\sigma_{\ddot{x}_1} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(j\omega\right)^2 \frac{k_1}{\left(j\omega\right)^2 m_1 + k_1}\left(-j\omega\right)^2 \frac{k_1}{\left(-j\omega\right)^2 m_1 + k_1} A_b v \frac{1}{j\omega}\frac{1}{\left(-j\omega\right)} d\omega} = \infty \tag{11.12}$$

In fact, for $\omega = \pm\sqrt{k_1/m_1}$, the integrand is unbounded. Also $\sigma_{F_z} = m_1\sigma_{\ddot{x}_1} = \infty$.

Summarizing the results of the aforementioned analyses, we can state that *one single suspension system is not efficient for filtering the excitation coming from uneven road profile* and two stages of suspension (like the ones represented in Figure 11.10d) are far more efficient, as experience suggests and as we will show as follows. This may explain why today's road vehicles are equipped with two suspension systems, one represented by the pneumatic tire and one represented by what is properly called the suspension system, that is, the system composed basically by a spring and a damper.

Let us now consider Figure 11.10d. Such a model may represent vehicles with different suspension configurations. In [2], it is shown how the model in Figure 11.10d can represent actual four-wheel vehicles. The performances that are relevant for studying the vertical vibrations are

- The body vertical acceleration (represented by $\ddot{x}_2$)
- The force at the wheel–road contact area (represented by F_z)
- The relative displacement wheel to body (represented by $x_2 - x_1$)

The discomfort of passengers carried by the vehicle in Figure 11.10d is evaluated by computing the standard deviation of the vertical vehicle body acceleration. The higher the standard deviation, the higher is the *discomfort*. This approach seems to provide very good correlation with subjective ride comfort ratings [68]. The standard deviation of the tire radial force is related to *road holding*. The variation of tire radial force can lead to a loss of contact with the ground and poor handling and braking ability [2,68]. The standard deviation of the relative displacement between the wheel and vehicle body, that is, the rattlespace or *working space*, is related to design and packaging constraints, as well as to wheel lateral vibrations. Discomfort, road holding, and working space are the *performance indices* to which reference will be made in the sequel.

The linear equations of motion of the system model in Figure 11.10d are

$$\begin{aligned} m_1\ddot{x}_1 - r_2\left(\dot{x}_2 - \dot{x}_1\right) - k_2\left(x_2 - x_1\right) + k_1 x_1 &= k_1\xi \\ m_2\ddot{x}_2 + r_2\left(\dot{x}_2 - \dot{x}_1\right) + k_2\left(x_2 - x_1\right) &= 0 \end{aligned} \tag{11.13}$$

The mass m_1 represents approximately the mass of the wheel plus part of the mass of the suspension arms, and m_2 represents approximately 1/4 of the body mass (m_2 could be computed more precisely as shown as follows). The damping parameter r_2 represents the first-order coefficient of the damper

characteristic, the stiffness parameter k_2 represents the first-order coefficient of the stiffness characteristic, and the stiffness parameter k_1 represents the tire radial stiffness at static loading. For more detail on the system model and its components, see [1,2,69,70].

The frequency response of the amplitudes $x_1 = X_1 e^{j\omega t}$, $x_2 = X_2 e^{j\omega t}$ due to the input $\xi = \Xi e^{j\omega t}$, $\Xi = 1$ is [68]

$$X_1(j\omega) = \frac{k_1\left(k_2 + jr_2\omega - m_2\omega^2\right)}{D(j\omega)} \tag{11.14}$$

$$X_2(j\omega) = \frac{k_1\left(k_2 + jr_2\omega\right)}{D(j\omega)} \tag{11.15}$$

$$\begin{aligned} D(j\omega) &= k_1k_2 + jk_1r_2\omega - \left(k_2m_1 + k_1m_2 + k_2m_2\right)\omega^2 \\ &\quad - jr_2\left(m_1 + m_2\right)\omega^3 + m_1m_2\omega^4 \end{aligned} \tag{11.16}$$

The frequency response of $\ddot{x}_2 = -\omega^2 X_2 e^{j\omega t}$ due to the input $\xi = \Xi e^{j\omega t}$, $\Xi = 1$ is [68]

$$H_1(j\omega) = -\omega^2 X_2(j\omega) \tag{11.17}$$

The frequency response of $F_z = \bar{F}_z e^{j\omega t}$ due to the input $\xi = \Xi e^{j\omega t}$, $\Xi = 1$ is [68]

$$H_2(j\omega) = k_1\left(1 - X_1(j\omega)\right) \tag{11.18}$$

The frequency response of $x_2 - x_1 = (X_2 - X_1)e^{j\omega t}$ due to the input $\xi = \Xi e^{j\omega t}$, $\Xi = 1$ is [68]

$$H_3(j\omega) = X_2(j\omega) - X_1(j\omega) \tag{11.19}$$

For $S_{\xi 1}$ excitation, the standard deviations of $\ddot{x}_2$, F_z, $x_2 - x_1$ are, respectively,

$$\sigma_{\ddot{x}_2} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|H_1\right|^2 S_{\xi 1}d\omega}, \quad \sigma_{F_z} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|H_2\right|^2 S_{\xi 1}d\omega}, \quad \sigma_{x_2 - x_1} = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|H_3\right|^2 S_{\xi 1}d\omega}$$

and, from [68],

$$\begin{aligned} \sigma_{\ddot{x}_2}^2 &= \frac{1}{2}A_b\upsilon\bar{\sigma}_{\ddot{x}_2}^2 \\ \bar{\sigma}_{\ddot{x}_2}^2 &= \frac{1}{m_2^2}\left(\frac{\left(m_2 + m_1\right)k_2^2}{r_2} + k_1r_2\right) \end{aligned} \tag{11.20}$$

$$\sigma_{F_z}^2 = \frac{1}{2} A_b \upsilon \bar{\sigma}_{F_z}^2$$

$$\bar{\sigma}_{F_z}^2 = \left(m_2 + m_1\right)^2 \left(P\right)$$

$$P = \left(\frac{\left(m_2 + m_1\right) k_2^2}{m_2^2 r_2} - \frac{2k_1 k_2 m_1}{m_2 r_2 \left(m_2 + m_1\right)} + \frac{k_1^2 m_1}{r_2 \left(m_2 + m_1\right)^2} + \frac{k_1 r_2}{m_2^2} \right) \tag{11.21}$$

$$\begin{aligned} \sigma_{x_2 - x_1}^2 &= \frac{1}{2} A_b \upsilon \bar{\sigma}_{x_2 - x_1}^2 \\ \bar{\sigma}_{x_2 - x_1}^2 &= \frac{\left(m_2 + m_1\right)}{r_2} \end{aligned} \tag{11.22}$$

For $S_{\xi 2}$ excitation, the standard deviations of $\ddot{x}_2$, F_z, $x_2 - x_1$ are, respectively,

$$\sigma_{\ddot{x}_2} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|H_1\right|^2 S_{\xi 2} d\omega}, \quad \sigma_{F_z} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|H_2\right|^2 S_{\xi 2} d\omega}, \quad \sigma_{x_2 - x_1} = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|H_3\right|^2 S_{\xi 2} d\omega}$$

and from [68]

$$\sigma_{\ddot{x}_2}^2 = c_{r\upsilon} \frac{\left[k_1^2 r_2^2 \left(k_2 + r_2 s_c + m_2 s_c^2\right) + k_1 k_2^2 \left(\left(m_1 + m_2\right) \left(k_2 + r_2 s_c\right) + m_1 m_2 s_c^2 \right) \right]}{\left(m_2^2 r_2 \left(D_S\right) \right)}$$

$$D_S = k_1 k_2 + k_1 r_2 s_c + k_2 \left(m_1 + m_2\right) s_c^2 + k_1 m_2 s_c^2 + \left(m_1 + m_2\right) r_2 s_c^3 + m_1 m_2 s_c^4$$

$$s_c = a\upsilon, \quad c_{r\upsilon} = \frac{1}{2 A_\upsilon a \upsilon} \tag{11.23}$$

$$\sigma_{F_z}^2 = c_{r\upsilon} \frac{\left(A_{2S} + B_{2S} + C_{2S}\right)}{k_1^2 m_2^2 r_2^2 \left(D_S\right)}$$

$$A_{2S} = k_1^4 \left(m_1 + m_2\right) r_2 \left(-2k_2 m_1 m_2 + m_1 r_2^2 + m_2 r_2^2\right) \left(k_2 + r_2 s_c + m_2 s_c^2\right)$$

$$B_{2s} = k_1^3 k_2^2 \left(m_1 + m_2\right)^2 r_2 \left(k_2 m_1 + k_2 m_2 + m_1 r_2 s_c + m_2 r_2 s_c + m_1 m_2 s_c^2\right)$$

$$C_{2s} = k_1^4 m_1 m_2^2 r_2 \left(k_1 k_2 + k_1 r_2 s_c + k_1 m_2 s_c^2 + k_2 m_2 s_c^2 + m_2 r_2 s_c^3\right) \tag{11.24}$$

$$\sigma_{x_2 - x_1}^2 = c_{r\upsilon} \frac{k_1 \left(k_2 m_1 + k_2 m_2 + m_1 r_2 s_c + m_2 r_2 s_c + m_1 m_2 s_c^2\right)}{r_2 \left(D_s\right)} \tag{11.25}$$

The formulae 11.20 through 11.25 state that the dynamic response of a vehicle model like the one in Figure 11.10d is *bounded*, due to either excitation $S_{\xi 1}$ or excitation $S_{\xi 2}$. The configuration in Figure 11.10d seems also the *simplest possible* to obtain a good filtered dynamic response for

a generic wheeled vehicle. This may explain why today's road vehicles are configured the way we commonly experience them, that is, a body connected by a suspension system to wheels with a built-in suspension system.

The engineering question arose before on the limit speed can be easily answered now by setting Equation 1.21 (or alternatively, Equation 11.24) equal to $p\sigma_{F_z} = F_{z,static} = (m_1 + m_2)g$ and then derive, for $p = 3$, the speed at which the wheel leaves the ground with 0.3% probability (obviously one may seek higher or lower probabilities, by changing properly the value of p).

The aforementioned reasoning shows that, for a wheeled vehicle running on a rough road profile, *a double-stage suspension system is necessary*. However, a word of caution is in order. The stochastic processes $S_{\xi 1}$ and $S_{\xi 2}$ are the idealization of real processes, and the aforementioned statements have to be taken as an engineering science contribution. For example, farm tractors are, still today, often unsprung, despite the acceleration on the body that is theoretically unbounded, yet they run mostly at vanishing speed, often on soft soil, and imperatively, the driver seat has always been sprung!

In Table 11.1, a fundamental sensitivity analysis is presented, based on Equations 11.20 through 11.22.

TABLE 11.1
Limits of the Three Respective Performance Indices $\sigma_{\ddot{x}_2}$, σ_{F_z}, $\sigma_{x_2-x_1}$ Computed by Equations 11.20 through 11.22, Respectively

	$\sigma_{\ddot{x}_2}$	σ_{F_z}	$\sigma_{x_2-x_1}$
$v \to 0$	0	0	0
$v \to \infty$	∞	∞	∞
$A_b \to 0$	0	0	0
$A_b \to \infty$	∞	∞	∞
$m_1 \to 0$	$\sqrt{\frac{A_b v}{2}\left[\frac{1}{m_2^2}\left(\frac{k_2^2 m_2}{r_2} + r_2 k_1\right)\right]}$	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{k_2^2 m_2}{r_2} + r_2 k_1\right)\right]}$	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_2}{r_2}\right)\right]}$
$m_1 \to \infty$	∞	∞	∞
$m_2 \to 0$	∞	∞	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_1}{r_2}\right)\right]}$
$m_2 \to \infty$	0	∞	∞
$r_2 \to 0$	∞	∞	∞
$r_2 \to \infty$	∞	∞	0
$k_1 \to 0$	$\sqrt{\frac{A_b v}{2}\left[\frac{1}{m_2^2}\left(\frac{k_2^2 (m_1 + m_2)}{r_2}\right)\right]}$	$\sqrt{\frac{A_b v}{2}\left[\frac{1}{m_2^2}\left(\frac{k_2^2 (m_1 + m_2)^3}{r_2}\right)\right]}$	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_1 + m_2}{r_2}\right)\right]}$
$k_1 \to \infty$	∞	∞	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_1 + m_2}{r_2}\right)\right]}$
$k_2 \to 0$	$\sqrt{\frac{A_b v}{2}\left[\frac{1}{m_2^2} k_1 r_2\right]}$	$\sqrt{\frac{A_b v}{2}\left[\frac{1}{m_2^2}\left(\frac{k_1 (k_1 m_1 m_2^2 + r_2^2 (m_1 + m_2)^2)}{r_2}\right)\right]}$	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_1 + m_2}{r_2}\right)\right]}$
$k_2 \to \infty$	∞	∞	$\sqrt{\frac{A_b v}{2}\left[\left(\frac{m_1 + m_2}{r_2}\right)\right]}$

Several consideration can be made by inspection of Table 11.1 and related Equations 11.20 through 11.22; $\sigma_{\ddot{x}_2}$, σ_{F_z}, $\sigma_{x_2-x_1}$ increase monotonically with v, A_b.
For what concerns $\sigma_{\ddot{x}_2}$, one may see that

- Reducing to zero the unsprung mass m_1 does not imply the vanishing of $\sigma_{\ddot{x}_2}$; an important increase of m_1 causes an important increase of $\sigma_{\ddot{x}_2}$.
- Reducing to zero the sprung mass m_2 implies very high $\sigma_{\ddot{x}_2}$; an important increase of m_2 causes the vanishing of $\sigma_{\ddot{x}_2}$.
- Reducing to zero the damping r_2 implies a strong increase of $\sigma_{\ddot{x}_2}$; increasing the damping r_2 implies a strong increase of $\sigma_{\ddot{x}_2}$; this means that there is an optimum damping that can be computed by setting $d\sigma_{\ddot{x}_2}/dr_2 = 0$, which gives

$$r_{2,opt} = \frac{k_2}{k_1}\sqrt{k_1\left(m_1 + m_2\right)}$$

In-depth analyses on the influence of the damping on the sprung mass accelerations have been given in the fundamental paper by Bourcier de Carbon [71] and in [2,67,68].

- Reducing to zero the tire stiffness k_1 does not imply the vanishing of $\sigma_{\ddot{x}_2}$ (this depends on the fact that the lower frequency mode becomes undamped as $k_1 \to 0$ and the frequency response function has a high peak near to the null frequency); on the contrary, an important increase of k_1 causes an important increase of $\sigma_{\ddot{x}_2}$; this means that an extremely strong reduction of tire stiffness is not beneficial for making $\sigma_{\ddot{x}_2}$ vanish.
- Reducing to zero the stiffness k_2 does not imply the vanishing of $\sigma_{\ddot{x}_2}$; an important increase of k_2 causes an important increase of $\sigma_{\ddot{x}_2}$; this means that an extremely strong reduction of suspension stiffness is not beneficial for making $\sigma_{\ddot{x}_2}$ vanish.

For what concerns σ_{F_z}, one may consider that $F_z = m_1\ddot{x}_1 + m_2\ddot{x}_2$, so the results are related to the ones already presented for $\sigma_{\ddot{x}_2}$; we can state that

- Reducing to zero the unsprung mass m_1 does not imply the vanishing of σ_{F_z}; on the contrary, an important increase of m_1 causes an important increase of σ_{F_z}.
- Reducing to zero the mass m_2 implies a strong increase of σ_{F_z}; increasing the mass m_2 implies a strong increase of σ_{F_z}; this means that there is an optimum m_2 that can be computed by setting $d\sigma_{F_z}/dm_2 = 0$, which gives involved analytical solutions that cannot be displayed here.
- Reducing to zero the damping r_2 implies a strong increase of σ_{F_z}; increasing the damping r_2 implies a strong increase of σ_{F_z}; this means that there is an optimum damping that can be computed by setting $d\sigma_{F_z}/dr_2 = 0$, which gives

$$r_{2,opt} = \frac{1}{k_1(m_1 + m_2)}\sqrt{k_1\left(k_2^2\left(m_1 + m_2\right)^3 + m_1m_2\left(m_2k_1^2 - 2m_2k_1k_2 - 2m_1k_1k_2\right)\right.}$$

- Reducing to zero the tire stiffness k_1 does not imply the vanishing of σ_{F_z}; on the contrary, an important increase of k_1 causes an important increase of σ_{F_z}; this means that an extremely strong reduction of tire stiffness is not beneficial for making σ_{F_z} vanish.
- Reducing to zero the stiffness k_2 does not imply the vanishing of σ_{F_z}; on the contrary, an important increase of k_2 causes an important increase of σ_{F_z}; this means that a strong increase of suspension stiffness is detrimental for a limited σ_{F_z}.

For what concerns $\sigma_{x_2-x_1}$, one may see that

- Reducing to zero the unsprung mass m_1 does not imply the vanishing of $\sigma_{x_2-x_1}$; on the contrary, an important increase of m_1 causes an important increase of $\sigma_{x_2-x_1}$.
- Reducing to zero the sprung mass m_2 does not imply the vanishing of $\sigma_{x_2-x_1}$; on the contrary, an important increase of m_2 causes an important increase of $\sigma_{x_2-x_1}$.
- Reducing to zero the damping r_2 implies a strong increase of $\sigma_{x_2-x_1}$; increasing the damping r_2 implies the vanishing of $\sigma_{x_2-x_1}$.
- $\sigma_{x_2-x_1}$ does not depend on k_1 and k_2, respectively.

All of the aforementioned considerations refer only to a fundamental parameter sensitivity analysis devoted to conceptual vehicle design considerations, referring to how many suspension stages have to be adopted for a ground vehicle. An in-depth sensitivity analysis referring to suspension system can be found in [2,68]. The optimal tuning of vehicle parameters is another additional question and does require much more theoretical concepts that are given in Chapter 8. Just one fundamental theoretical issue on optimal suspension parameter setting is highlighted here. By comparing Equation 11.20 with Equation 11.23 and Equation 11.21 with Equation 11.24 and Equation 11.22 with Equation 11.25, it follows that if the excitation is $S_{\xi 1}$, then the optimal suspension parameter setting is independent on the vehicle speed. Actually, in Equations 11.20 and 11.21, A_b and v are not mixed with the parameters m_1, m_2, k_1, k_2, r_2. On the contrary, for excitation $S_{\xi 2}$, the parameters A_v, s_c are mixed with the parameters m_1, m_2, k_1, k_2, r_2. The performance computed by means of Equations 11.20 through 11.22 is not much different with respect to the ones computed by means of Equations 11.23 through 11.25 [68] (we leave to the reader the proper analysis). From a theoretical point of view, the relevant point is that with $S_{\xi 1}$ excitation, the suspension tuning is independent from vehicle speed; on the contrary, with $S_{\xi 2}$ excitation, the suspension tuning is dependent from vehicle speed, which is an important issue for vehicle architecture.

Summarizing the issues of this section, we have stated that a passive suspension system is necessary only if a vehicle have to move at relatively high speed. *If vehicles were running on a perfectly smooth surface, no suspension system would be necessary.* No road is smooth enough to avoid excitation of vibration in the vertical direction, unless the vehicle speed is vanishing. Double-stage suspensions were (and are) adopted for vehicles to cope with the vibrations induced by road irregularity.

The aforementioned considerations have been referred to passive suspension systems. The results hold for adaptive suspensions. The results do not hold for ideal active suspensions (see Chapter 31).

Referring to passive suspensions, the question arises on how setting the parameters of the suspension systems in Figure 11.10d, in order to obtain the best performance in terms of minimum $\sigma_{\ddot{x}_2}$, minimum σ_{F_z}, and minimum $\sigma_{x_2-x_1}$. In [68] and in Chapter 8, the question is answered.

The aforementioned considerations were made referring to a randomly uneven profile. In a real environment, excitation is produced by either continuous or isolated obstacles or a combination of them. Taking into account such isolated obstacles (curbs, potholes, etc.), the main general results would have been the same. The description of isolated obstacles and how modeling a pneumatic wheel hitting them is given in [2,7].

In [2], the addition of a spring in series to the damper is discussed with reference to the influence on vibration filtering.

11.3.2.1 Lumped Mass Car Body

The considerations made earlier refer to a so-called quarter-car model [1,2,7,69]. Actually, the quarter-car model is often a sufficient approximation of a full-car model [2] when studying vertical

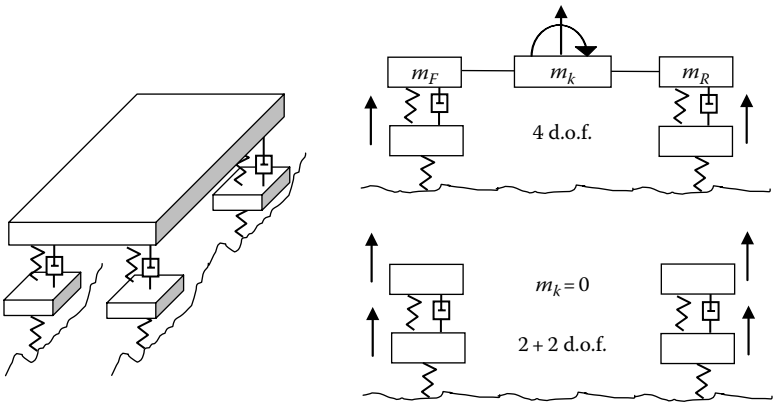


FIGURE 11.11 Lumped mass car models.

vibrations. The unsprung mass m_2 in Figure 11.10d can be determined after considering the half-car model depicted in Figure 11.11; its mass can be partitioned in three parts, such that

$$m_F + m_R + m_k = M$$

$$J = m_F l_F^2 + m_R l_R^2$$

where

M is the mass of the car body
J is the pitch moment of inertia around the center of gravity
m_k is the coupling mass at the center of gravity
m_F, m_R are the sprung masses of the fore and rear quarter models in Figure 11.11
l_F and l_R are the respective horizontal distances of the center of gravity from the front and rear axles

Introducing the condition of linear momentum $0 = m_F l_F - m_R l_R$, the three unknowns m_F, m_R, m_k can be found and read

$$m_F = \frac{J}{(l_F + l_R) l_F}$$

$$m_R = \frac{J}{(l_F + l_R) l_R} \tag{11.26}$$

$$m_k = \frac{M l_F l_R - J}{l_F l_R}$$

If $J = M l_F l_R$, then $m_k = 0$, that is, the front and rear parts of the half model are *decoupled* and $m_F = M l_R / l$, $m_R = M l_F / l$ where l is the wheelbase, $l = l_F + l_R$. If $J = 0$, then $m_k = M$. Often $m_k \approx 0$ in actual cars, so quarter-car models allow to investigate the dynamics of vehicles with the minimum number of variables and parameters.

In this reasoning, roll motion has been disregarded; the complete treatment on how a quarter-car model can be derived is given in [2]; the conclusion is that the reasoning presented earlier can be adopted often with reasonable accuracy for conceptual design investigations.

11.3.2.2 Ride Comfort and Road Holding Influenced by Main Vehicle Dimensions

The wheelbase and the track dimensions influence the performance of the vehicle while running on a road with a random unevenness.

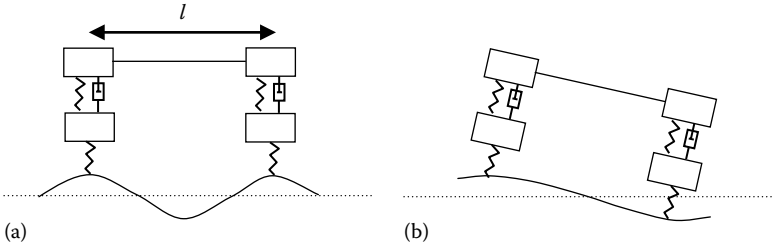


FIGURE 11.12 Heave and pitch motions. (a) Heave excited only and (b) pitch excited only.

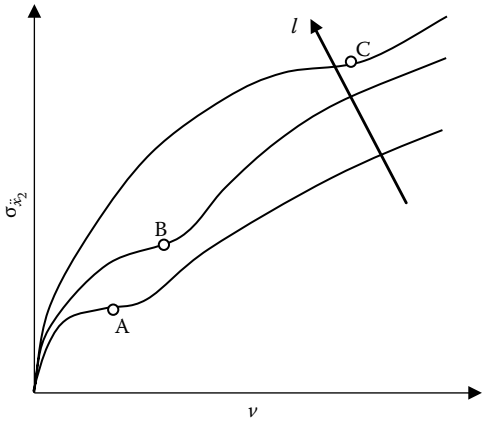


FIGURE 11.13 Typical ride discomfort of an automobile as function of wheelbase l and vehicle speed v.

The wheelbase l influences the ride comfort as, depending on the wavelength of the road unevenness, the heave or the pitch can be or cannot be excited; a simple explanation of this fact can be argued by inspection of Figure 11.12 [2]. Such an occurrence is called *wheelbase filtering*.

Due to wheelbase filtering, the ride discomfort as defined in Section 11.3.2 (and in a more general way in [2)]) depends on vehicle speed v and wheelbase l as shown in Figure 11.13. If a given speed would be attained more frequently than other ones, the wheelbase could be set in order to exploit the bends A, B, or C; more details are reported in [2].

Referring to Figure 11.9, the position of the seat (L114) is of crucial importance for the ride comfort of the passenger, as shown in Figure 11.14 [2]. The seat nearby the center of gravity (mid-wheelbase) is better than a seat on an axle; details on how computing the ride comfort are given in [1,2,66] and in Chapters 29 and 36.

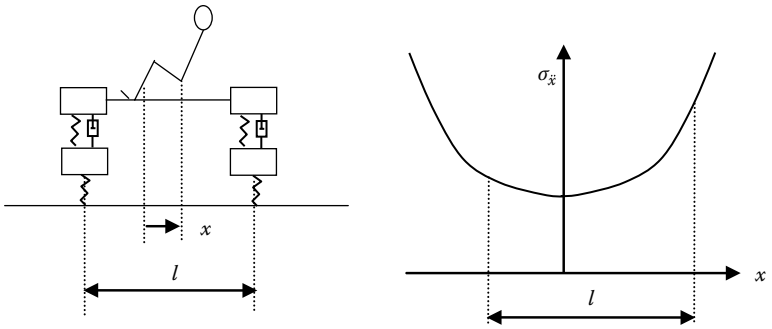


FIGURE 11.14 Ride discomfort as function of the position of the seat along the wheelbase.

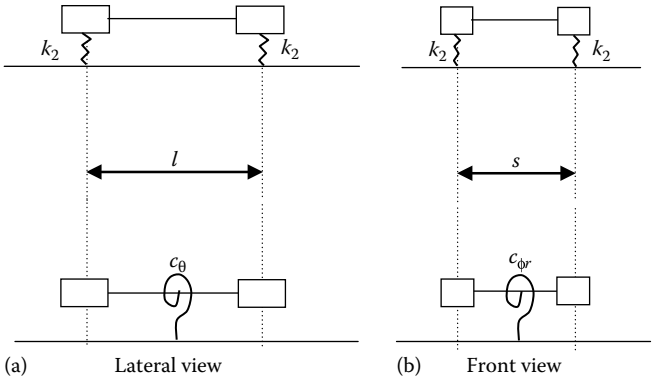


FIGURE 11.15 (a) Anti-dive effect of the wheelbase l and (b) antiroll effect of the track s.

A given amount of pitch stiffness is required to avoid car body dive during braking. The complete and comprehensive theory of anti-dive and anti-squat is given in [2,72]; see Chapter 21. Referring to Figure 11.15, the pitch stiffness is $C_\theta = k_2 l^2/2$, that is, the pitch stiffness depends on the square of the wheelbase l. At a given pitch stiffness, a slight increase of the wheelbase allows a substantial decrease of the suspension stiffness, thus improving ride comfort (which is positively affected by a low k_2).

The road holding is not influenced directly by the wheelbase, at least if $m_k = 0$, which is a reasonable assumption for automobiles.

The total roll stiffness $c_\phi = c_\phi + c_{\phi b}$ depends on the square of the track s, where $c_{\phi r}$ is related to the effect of the suspension springs $c_{\phi r} = k_2 s^2/2$ (such an effect is similar to the one referring to pitch; Figure 11.15) and $c_{\phi b}$ is related to the effect of antiroll bars (see [1,2,7]) $c_{\phi b} = k_b s^2/b^2$ (see Figures 11.15 and 11.16). At a given roll stiffness, a slight increase of the track allows a substantial decrease of the suspension stiffness, thus improving ride comfort (which is positively affected by a low k_2).

In [2], the coherence function γ is effectively introduced to relate the power spectral density of the random vertical unevenness at the track center (see Equations 11.1 and 11.2) $(1/2)(1 + \gamma)S_\xi i(\omega)$ $(i = 1, 2)$, with the power spectral density of roll unevenness (antisymmetrical unevenness at the left and right wheel), and $(2/s^2)(1 - \gamma)S_{\xi i}(\omega)$ $(i = 1, 2)$. The last power spectral density states that the excitation of the roll motions is *inversely proportional to the square of the track*.

11.3.3 Pneumatic Wheel

We have seen that *a double-stage suspension system is necessary* (if the common technology is exploited) and that the first suspension stage can be simply a spring. We dare state that the first suspension stage *must* be simply a spring as a mass in contact with the ground would be subject to unbounded inertia forces; actually the system in Figure 11.17 is the sum of the model in Figure 11.10a plus the model in Figure 11.10d (for the sake of space, the proof of this statement is omitted here).

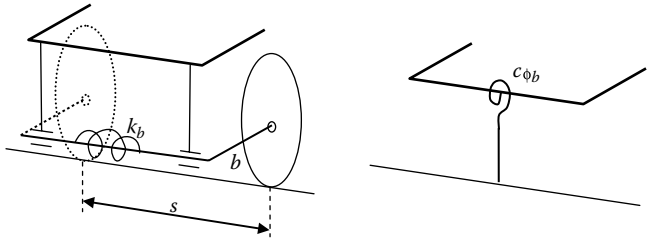


FIGURE 11.16 Antiroll bar. Torsion beam stiffness k_b and equivalent roll stiffness $c_{\phi b}$ (compare with Figure 11.15b).

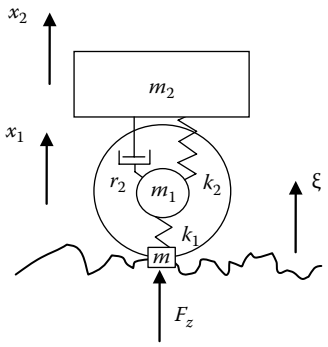


FIGURE 11.17 Model depicted in Figure 11.10d, with the additional mass m.

The pneumatic tire can be modeled [1,2,69] by means of a massless spring, as Figure 11.10d shows. It is the well-known solution referring to Thomson and Dunlop inventions [9].

The main functions of the pneumatic wheel are four; in addition to cushioning (i.e., the first stage of the suspension system), the other three functions are simply carrying the load F_z, applying to the ground the traction/braking forces F_x and applying to the ground the cornering forces F_y.

11.3.3.1 Load Carrying and Radial Stiffness

The load-carrying function of the pneumatic wheel is in principle fulfilled as shown in Figure 11.18 where a torus is pressed against a hard ground. If the torus is made from a membrane (unable to resist flexural moments), the part of the torus in contact with the ground is loaded, respectively, in vertical direction by the double opposing pressures p_i and, in horizontal direction, by distributed tension t. Considering the equilibrium in the vertical direction, the vertical force is $F_z = p_i A_c$ and the pressure on the ground is equal to the inner pressure p_i. The ideal pneumatic tire of Figure 11.18 *is not subject to the Hertz contact pressure*; actually the contact pressure is *constant*. In such an ideal case, the load-carrying capacity of the pneumatic tire is given only by the inner pressure. Unfortunately, the compliance of the torus would be very weak in the horizontal direction, that is, the deformations associated to forces F_x, F_y would be too high (Figure 11.18). So a kind of a stiff ring at the outer portion of the torus has been conceived and applied as shown in Figure 11.19 where a cross section of a radial tire is shown [8,10,73,74] (in the sequel, a "radial tire"—no "run-flat" type—will be considered; older types of pneumatic tires (bias) behave similarly to a stiffened torus, approaching the behavior of radial tires).

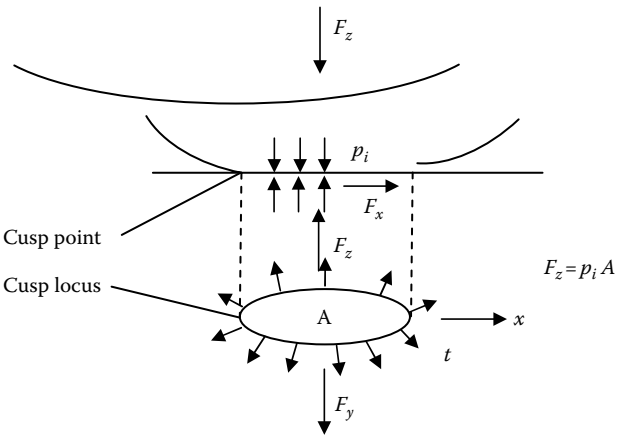


FIGURE 11.18 Membrane torus pressed on hard ground.

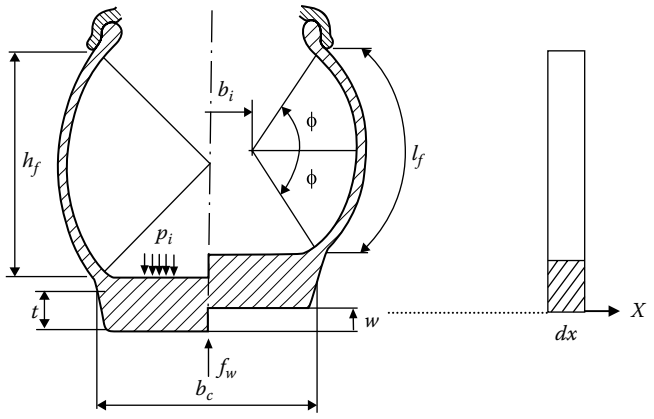


FIGURE 11.19 Cross section of a pneumatic tire (radial type). (Adapted from Mastinu, G. and Fainello, M., Study of the pneumatic tyre behaviour on dry and rigid road by finite element method, *Vehicle System Dynamics*, 21, 143, 1992, http://www.informaworld.com/smpp/title~db=all~content=t713659010~tab=issueslist~branches=27-v271.)

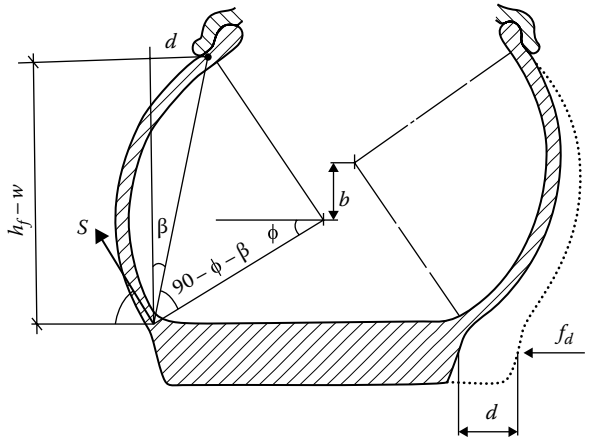


FIGURE 11.20 Section of a pneumatic tire deflected laterally. (From Mastinu, G. and Fainello, M., Study of the pneumatic tyre behaviour on dry and rigid road by finite element method, *Vehicle System Dynamics*, 21, 143, 1992, http://www.informaworld.com/smpp/title~db=all~content=t713659010~tab=issueslist~branches=27-v271.)

Let us consider Figures 11.19 and 11.20 [109]. The respective meanings of the introduced symbols are as follows:

h_f is the height of the sidewall
b_c is the cross width of the belt
w is the vertical (radial) displacement
l_f is the length of the sidewall
f_w is the radial force per unit length (in radial direction)
f_d is the lateral force per unit length (in x direction, i.e., circumferential direction)
p_i is the inner pressure

By inspection of Figure 11.21, the equilibrium in vertical direction of the belt is given by (forces are given per unit length dx along the direction of motion of the wheel)

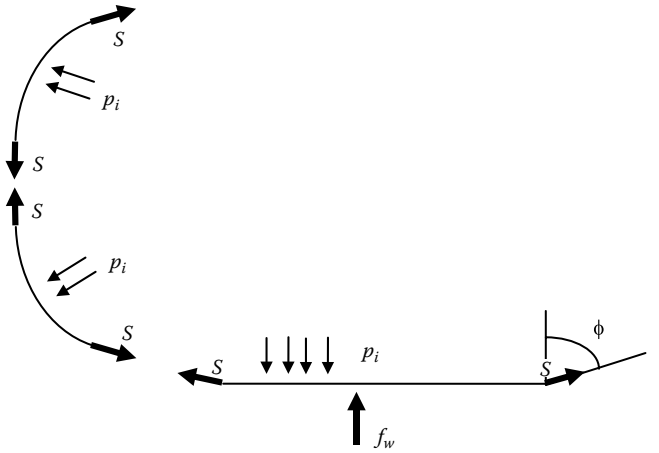


FIGURE 11.21 Cross section of a pneumatic tire with internal and external forces.

$$2S\cos(\phi) - p_i b_c + f_w = 0 \quad \text{additionally}$$

$$\tan(\phi) = \frac{(h_f - w)/2}{(b_c - b_i)/2} = \frac{S\sin(\phi)}{S\cos(\phi)} = \frac{p_i(h_f - w)/2}{S\cos(\phi)}$$

Thus, $S\cos(\phi) = p_i(b_c - b_i)/2$, which implies

$$f_w = p_i b_i \tag{11.27}$$

In order to compute f_d, for small displacements, let us consider that $\tan(\beta) = \dfrac{d}{h_f - w} \cong \sin(\beta) \cong \beta$ and that

$$S_{ys} = S\cos(90 - \phi + \beta)$$

$$S_{yd} = S\cos(90 - \phi - \beta)$$

$$f_d = S_{ys} - S_{yd}$$

Remembering that $S\cos(\phi) = p_i(b_c - b_i)/2$, and using the known trigonometric transformation rules, we can write

$$f_d = 2S\cos(\phi)\sin\frac{d}{h_f - w} = p_i\frac{b_c - b_i}{h_f - w}d \tag{11.28}$$

From (11.27), if $f_w = 0$, then $b_i = w = 0$, and $f_d = p_i(b_c/h_f)d$, which states that the lateral stiffness per unit length (along the circumference direction) of a pneumatic tire (f_d/d) is higher if the *aspect ratio* $a_r = h_f/b_c$ is lower. This explains why the torus-like tire (whose aspect ratio $b_c/h_f = 1$) has been soon abandoned for stiffer pneumatic tires in lateral direction, with lower aspect ratio (a_r = 0.70–30). Less stiff tires in lateral direction have the so-called relaxation length too high that implies a lazy response of the tire in producing lateral forces due to steering input (see [1,2,8,10]).

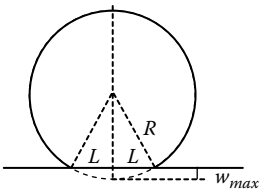


FIGURE 11.22 Approximated contact patch length of a pneumatic tire model whose cross section is shown in Figure 11.19.

Equation 11.27 states that the load-carrying capacity of the tire is related not only to *inner pressure* but also to *geometry*. The pressure does not increase substantially with the increasing load (due to the very limited variation of volume). By inspection of Figure 11.22, the vertical force F_z could be computed summing all of the contributions $f_w(x) = p_i b_i(x)$ at a different coordinate x along the contact patch [10,75]:

$$F_z = \int_{-L}^{L} f_w(x)\,dx = p_i \int_{-L}^{L} b_i(x)\,dx \tag{11.29}$$

The higher the pressure, the higher the load that can be carried.

An analytical formula for Equation 11.29 is reported in [76]:

$$\begin{aligned}
F_z &= 2p_i\left(A\left(w_{max}\right)+B\left(w_{max}\right)+C\left(w_{max}\right)\right)+k_{carc}w_{max}\\
A &= \sqrt{2Rw_{max}}\left(b_c-b_{v0}\right)\\
B &= \frac{1}{2}\sqrt{\frac{1}{3}Ra_rk_{ar}b_c}\left(-\frac{\pi}{2}-\arctan\left(\frac{w_{max}+a_rb_c-a_rk_{ar}b_c}{2\sqrt{w_{max}}\sqrt{a_rk_{ar}b_c-a_rb_c}}\right)\right)a_rk_{ar}b_c\\
C &= \frac{1}{4}\sqrt{3Ra_rk_{ar}b_c}\\
&\quad\times\left(\left(a_rk_{ar}b_c-a_rb_c+w_{max}\right)\left(\frac{\pi}{2}+\arctan\left(\frac{w_{max}+a_rb_c-a_rk_{ar}b_c}{2\sqrt{w_{max}}\sqrt{a_rk_{ar}b_c-a_rb_c}}\right)\right)+2\sqrt{w_{max}}\sqrt{a_rk_{ar}b_c-a_rb_c}\right)
\end{aligned} \tag{11.30}$$

where the respective meanings of the introduced symbols are

R is the outer radius (half of the overall diameter)
$a_r = h_f/b_c$ is the aspect ratio
$a_f = a_rk_{ar} = l_f/b_c$
w_{max} is the maximum tire deflection
b_{v0} is the geometric parameter related to the hoop tension inside the belt
k_{carc} is the radial stiffness of the carcass when $p_i = 0$

Typical tire data are reported in Table 11.2.

A is roughly half of the contact area $A_c = 2A = 2b_cL$ at the tire–ground interface, so $F_{zi} = 2p_ib_cL$ represents the vertical force that would be carried by an ideal tire made from a membrane (see Figure 11.18) whose contact area was rectangular. The force F_{zi} is very high, so other

TABLE 11.2
Typical Tire Data of Equation 11.30, Internal Pressure p_i = 2 Bar

	165/70R14	205/55R16	265/50R19
l_f (mm)	97.8	89	113.5
b_c (measured) (mm)	134.4	177.8	230.5
h_f (mm)	88.7	79.8	103.5
b_{v0} (mm)	41.2	100.3	116.4
R (measured) (mm)	294.3	317.3	379.9
k_{carc} (kN/m)	36.1	51.6	61.9

terms (*B, C*, etc.) reduce such a high force to a value close to the real one F_z. The terms *B* and *C* refer to the forces *S* exerted by sidewalls as shown in Figure 11.21, such *S* forces act in order to reduce F_{zi} (see Figure 11.21).

By inspection (and analysis) of Equation 11.30, the *load-carrying capacity of a tire appears to be proportional to the inner pressure p_i and approximately proportional to the cross width b_c and to $\sqrt{R}$.*

This means that given the static load $F_{z,stat}$, the dimensions of the wheel will have approximately to be proportional to the product $p_i b_c \sqrt{2R}$ (see Equation 11.30). For resisting lateral forces, additional reasoning is needed, which is given in Section 11.4.1.1.

Another formula (referring to old tires) for the tire load-carrying capacity is reported in [18].

The radial stiffness at static condition can be computed from Equation 11.30 taking the derivative $\partial F_z/\partial w_{max}$ at $w_{max} = w_{max,stat}$ (i.e., at the statical w_{max}). The radial stiffness is *proportional to the inner pressure p_i and, with very good approximation, proportional to the cross width b_c and to $\sqrt{R}$.*

This is almost consistent with the formula reported in [18,77] for tire stiffness K_z (kg/mm):

$$K_z = 0.00028\,P\sqrt{OD\,W} + 3.45$$

where

P is the tire inflation pressure (kPa)
W is the tire footprint width (mm)
OD is the outside diameter (mm)

Given the load to be carried, the combination pressure–diameter is defined by the tire manufacturer relying on many aspects (safety, durability) that cannot be dealt with here [10,18]. The basic mathematical relation between the load-carrying ability, the inner pressure, and the cross width is suggested by Equation 11.30. Obviously, the contact of the tire with the ground is much more complicated than addressed here (see, e.g., Chapters 17 through 19 [8–10,78,79,110]).

11.3.3.2 Rolling Resistance

Tire rolling resistance is dealt with in [10] and particularly in [18]. In this section, we will give basic concepts arguing that rolling resistance is related to the tire elasticity necessary for running on non-smooth road. There are several phenomena that more or less qualitatively explain such resistance. Some simple phenomena are presented hereafter (for other phenomena, see "Tire Dynamics" chapter in [8] and [10]).

One simple phenomenon for rolling resistance can be explained qualitatively by means of the model in Figure 11.23a. In the contact patch, the radial force *F* starts from zero at the leading edge and comes back to zero at the trailing edge. The force *F* is directly related to the displacement in radial direction *s* of the generic element of the tire that is in contact with the ground; *s* (together with *F*) is null at both

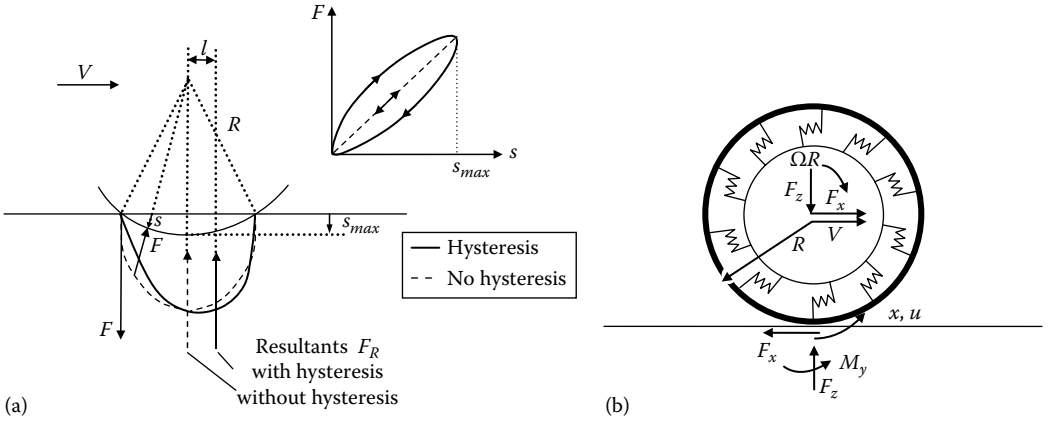


FIGURE 11.23 Tire models suitable for describing basic phenomena associated to rolling resistance: (a) hysteresis in radial direction, (b) hysteresis in circumferential direction.

the leading and trailing edges. There is a linear relationship $F = F(s)$ for hysteresis-free material. If the loading phenomenon is free from hysteresis, the resulting shape of the pressure at the contact area is symmetrical and the resultant force is applied at the center of the contact area. Otherwise, if the loading phenomenon is not hysteresis-free, then the resultant shape of the pressure at the contact area is *not* symmetrical and the resultant force is displaced from the contact area center toward its leading edge. This causes a moment $F_R l$ opposing the rotation of the wheel.

Another mechanism for describing rolling resistance is related simply to rolling. Let us consider Figure 11.23b where a curved bar of small (infinitesimal) thickness in radial direction is connected via a distributed stiffness k to a rigid disk. The curved reference coordinate x is attached to the disk whose center runs at constant forward speed V. The local displacement $u = u(x)$ of the bar is due to the force F_x that is applied to the only contact point between the tire and ground. Due to equilibrium, a moment $M_y = F_x R$ *must* be applied at the tire–ground contact point.

Let us imagine that the deformed part of the bar is located nearby the contact point. Let us write the equilibrium for a section dx of the bar

$$N - N - dN - kdxu = 0, \quad \text{that is, } \frac{dN}{dx} = -ku$$

where N is the compression force inside the bar. Hook's law states that $\sigma = N/A = E(-du/dx)$, where A is the cross section of the curved bar and E is Young's modulus (we have assumed positive compression force N, so the deformation ε reads $\varepsilon = -du/dx$). Taking the first derivative of Hook's law $dN/dx = -EA(d^2u/dx^2)$ and remembering that $dN/dx = -ku$, the differential equation governing the displacement $u = u(x)$ is derived

$$EA\frac{d^2u}{dx^2} - ku = 0$$

The solution of such a differential equation is $u = u(x) = U_0 e^{\lambda x}$. Substituting $u = U_0 e^{\lambda x}$ into the differential equation, the expression for λ can be derived, that is,

$$\lambda = \pm\sqrt{\frac{k}{EA}}$$

The boundary conditions are

$u = 0 \ @ \ x \to \pi R \sim \infty$

$u = 0 \ @ \ x \to -\pi R \sim -\infty$

thus,

$u = u(x) = U_0 e^{|\lambda| x} \ (x < 0)$

$u = u(x) = U_0 e^{-|\lambda| x} \ (x > 0)$

Additionally, at the contact point between the wheel and the ground, the contact force F_x is divided into two equal parts acting at the two bar tips (see Figure 11.24). At $x = 0^+$ ($x > 0$), $N = -F_x/2$, so applying Hook's law at such a bar tip and remembering that $u = u(x) = U_0 e^{-|\lambda| x}$,

$$-\frac{F_x}{2} = EA\left(-\frac{du}{dx}\right) = EA\left(-(-|\lambda|)U_0 e^{|\lambda|0}\right) = EA\lambda U_0 \quad \text{with} \quad \lambda = \sqrt{\frac{k}{EA}}$$

and so

$$U_0 = -\frac{F_x}{2EA\lambda}$$

that is, the bar has a negative displacement U_0 at the contact point. U_0 is "small," that is, the two points O and A in Figure 11.24 are separated by an infinitesimal distance.

Let us imagine that the rigid movement of the disk occurs around an axis through point O′ (Figure 11.25). As the curved bar can be deformed in circumferential direction, and point B must remain stuck at point O to the ground, the bar is elongated toward left by the amount *du*. In other words, during the motion, due to the bar elasticity, point B will remain at point O and will not reach the position shown in Figure 11.25.

The absolute velocity of point O with respect to the ground is $\vec{v}_{abs} = 0$; the relative velocity of point O (belonging to the curved bar) with respect to the disk is $\vec{v}_{rel} = du/dt < 0$ (*du* points in opposite

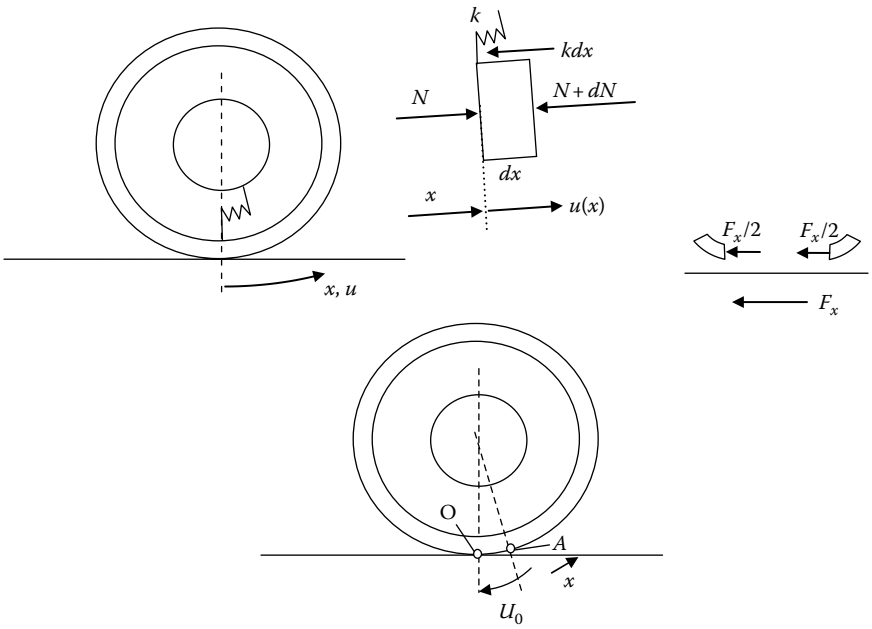


FIGURE 11.24 Tire model for describing the basic features of a non-stiff rolling wheel (see also Figure 11.23b).

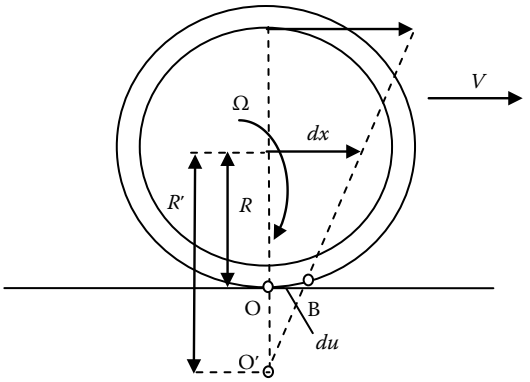


FIGURE 11.25 Rolling of a non-stiff wheel (see also Figure 11.23b). All displacements are small (infinitesimal).

direction with respect to *x*), and the velocity of the point O as it was attached to the disk is $\vec{v}_t$ whose projection along the *x*-axis is $\Omega(R' - R) = V - \Omega R$.

Let us consider that at point O,

$$\frac{du}{dt} = \left.\frac{du}{dx}\right|_{x=-U_0} \frac{dx}{dt}$$

and (remembering that $u = u(x) = U_0 e^{|\lambda| x}$ for $x < 0$) with $\lambda = \sqrt{k/EA}$
$du/dt = \lambda U_0 e^{\lambda U_0} \cdot V \cong \lambda U_0 V$ (as $dx/dt = V$ and $e^{\lambda U_0} \cong 1$ being $U_0 \cong 0$); thus,

$$\vec{v}_{rel} = \frac{du}{dt} = \lambda\left(\frac{-F_x}{2EA\lambda}\right)V = \frac{-F_x}{2EA}V < 0$$

At point O,

$$\vec{v}_{abs} = \vec{v}_{rel} + \vec{v}_t = \frac{-F_x}{2EA}V + V - \Omega R = 0$$

and rearranging,

$$F_x = 2EA\frac{V - \Omega R}{V}$$

The last expression states that the force F_x depends on both EA and the longitudinal slip $s_x = V - \Omega R/V$, that is, the longitudinal force can only develop if slip occurs. Since $F_x = M_y R$ (Figure 11.23b), if M_y is null, F_x is null and the wheel rolls at zero slip (ideal case). If hysteresis is present in the elastic bar, some power is needed to make the wheel rolling. In this case, neither M_y nor F_x is null and slip occurs. The rolling resistance is related to how much material is present in the elements that are deformed during rolling; hysteresis power is proportional to the volume of the deformed material. The cross area A can be expressed as $A = b_c t$ where b_c is the cross width of the tread (Figure 11.19) and t is the tread height (Figure 11.19). If E were extremely high, the deformations would be vanishing and the rolling resistance would be vanishing. *Inherently, not perfectly stiff wheels are prone to rolling resistance*. Material hysteresis is directly related to the rolling resistance. The displacements of the bar and related deformations are reported in Figure 11.26.

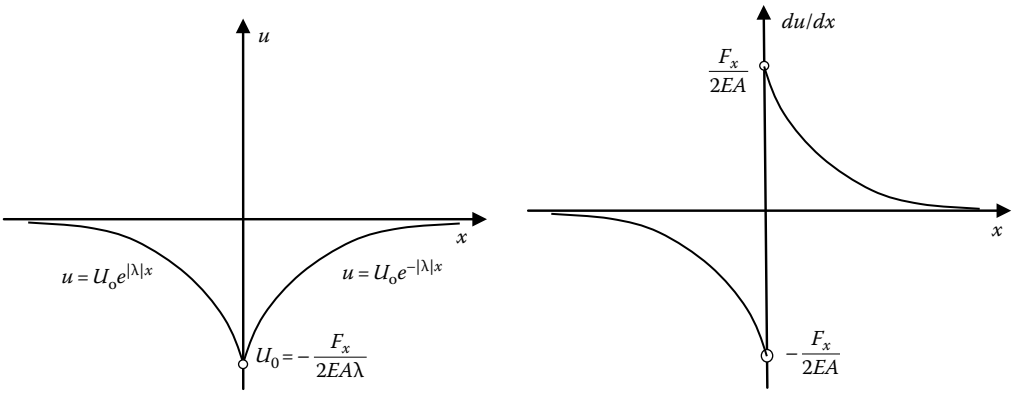


FIGURE 11.26 Displacement (u) and deformation (du/dx) of the elastic curved bar (wheel rim).

11.3.4 Powertrain Concept

In the preceding sections, we have seen that an elastic wheel is necessary for a vehicle to run on an uneven road and that the hysteresis associated with such an elasticity causes wheel rolling resistance. In this section, we will deal with the energy required to travel and the related impact on the concept of the powertrain.

In Chapter 27 Section 27.2, the longitudinal dynamic behavior of vehicles has been presented. In particular, driving resistances have been dealt with referring to straight driving.* Starting from such a modeling of vehicle dynamics, the conceptual design of the powertrain has to take into account the actual operation conditions, that is, the *mission* of the vehicle.

The reference cycle (criticized) describing the mission of European vehicles is the *New European Driving Cycle* (NEDC, also denoted as MVEG-95, Motor Vehicle Emission Group-95; Figure 11.27). In the United States, the corresponding cycle FTP-75 is used. For cars, design and construction solutions depend directly from the considered reference cycle.

In [80], a contribution to the conceptual design of powertrains is given. Taking into account the NEDC, the energy E required to travel 100 km is computed by considering the actual energy spent for running at different speeds:

$$E = A \cdot C_x \cdot 1.9 \times 10^4 + m \cdot f_R \cdot 8.4 \times 10^2 + m \cdot 10 \text{ kJ/100 km} \tag{11.31}$$

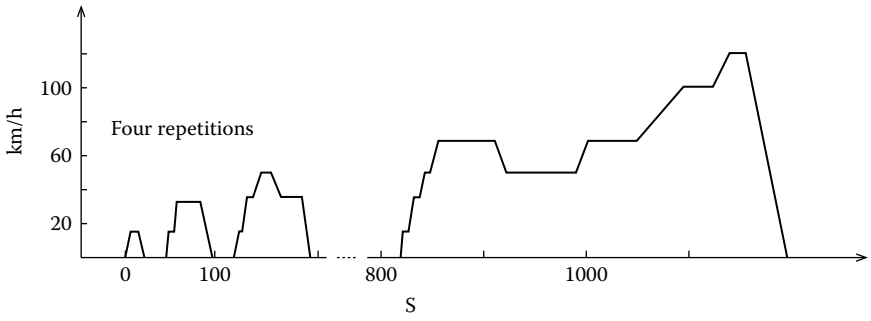
where

A is the cross-sectional area of the vehicle
C_x is the drag coefficient
m is the vehicle mass
f_R is the rolling resistance of tires

Introducing the proper parameter values in Equation 11.31, it appears that all of the three terms of the sum are of the same order of magnitude.

---

* During cornering, an additional rolling resistance appears due to the components of the lateral forces in the direction of forward motion of the car. Such a resistance is $F_{rx} = \frac{mv_x^2}{Rl}(l_F\alpha_R + l_R\alpha_F)$; the meaning of the symbols can be argued by inspection of Figure 11.31. Since the lateral slip at the axle is proportional to the lateral force at the same axle, that is, $\alpha_i \propto F_{yi}$ ($i = F, R$), and since $F_{yi} \alpha v_x^2$, it follows that $F_{rx} \alpha v_x^4$! The only way to reduce rolling resistance during cornering is to have low α_i, that is, high cornering stiffness.



	Unit	ECE 15	EUDC
Distance	[km]	4 × 1.013 = 4.052	6.955
Duration	[s]	4 × 195 = 780	400
Average speed	[km/h]	18.7 (with idling)	62.6
Maximum speed	[km/h]	50	120

FIGURE 11.27 NEDC (or MVEG-95), composed by four repetitions of ECE 15 driving cycle and one repetition of EUDC. (From Guzzella, L. and Sciarretta, A., *Vehicle Propulsion Systems*, Springer Verlag, Berlin, Germany, 2005, ISBN-10 3-540-25195-2.)

The relationship between the cross-sectional area A and the mass of today's cars can be taken as linear. The linear interpolation of the available data gives $A = a{\cdot}m + b$, where a and b (see [2]) are constants ($a = 5.7 \times 10^{-4}$ m^2/kg and $b = 1.172$ m^2).

The sensitivity of E with respect to the aerodynamic drag coefficient $p_1 = C_x$, to the tire rolling resistance $p_2 = f_R$, and to the vehicle mass $p_3 = m$ can be defined as

$$\lim_{\delta p_i \to 0} \frac{\left[E\left(p_i + \delta p_i\right) - E\left(p_i\right)\right]/E\left(p_i\right)}{\delta p_i / p_i} = \frac{\partial E}{\delta p_i}\frac{p_i}{E} \quad i = 1,3 \tag{11.32}$$

where

$$\frac{\partial E}{\delta p_1} = \frac{\partial E}{\partial C_x} = A \cdot 1.9 \times 10^4 \text{ kJ/100 km} \cdot \text{m}^2$$

$$\frac{\partial E}{\delta p_2} = \frac{\partial E}{\partial f_R} = m \cdot 8.4 \times 10^2 \text{ kJ/100 km}$$

$$\frac{\partial E}{\delta p_3} = \frac{\partial E}{\partial m} = a \cdot C_x \cdot 1.9 \times 10^4 + f_R \cdot 8.4 \times 10^2 + 10 \text{ kJ/100 km} \cdot \text{kg}$$

Expression (11.32) defines the normalized variation of E due to a unitary normalized variation of p_i. In Table 11.3, the E percent variations for 10% variation of p_i are reported for actual vehicles of different sizes. It is immediately evident that the highest sensitivity is related to mass. Reducing the vehicle mass by 10% can reduce by slightly less than 10% the energy required to travel 100 km at the different speeds defined by NEDC.

In Section 11.5, the mass distribution of different cars is introduced. By inspection of Table 11.3, it appears that reducing vehicle mass by a certain percent amount means reducing the mass of almost every subsystem by the same percent amount.

TABLE 11.3
Variation of E (Equation 11.32) due to Single Parameter Variation (for Different Vehicles)

	Data					Percent Variation of E due to 10% Variation Of		
Vehicle Type	Rated power (kW)	A (m²)	C_x	f_R	m (kg)	Aerodynamic drag coeff. $p_1 = C_x$	Rolling resistance $p_2 = f_R$	Vehicle mass $p_3 = m$
Midrange	140	2.2	0.26	12×10^{-3}	1560	2	4	9
Compact	55	2.0	0.29	10×10^{-3}	1120	4	3	8
Sports	310	1.95	0.29	12×10^{-3}	1650	2	4	9
SUV	200	2.3	0.41	14×10^{-3}	2640	2	4	9

Vehicle mass reduction is a more effective way to reduce the energy required to travel. Thus, mass reduction is a need for green mobility and a target (challenge) for the present and future of automotive technology.

Mass reduction allows vehicles to accelerate more rapidly, which is an important customer demand. For cars with internal combustion engine, an approximated formula for obtaining the acceleration time t_{vx} to reach the speed v_x from still is $t_{v_x} \cong v_x^2 \cdot m/P_{max}$ where P_{max} is the maximum rated power of the engine and m the vehicle mass.

In Table 11.4, referring to different vehicles, the ratio $E/\Delta T$ for completing the NEDC is reported, where E has been defined before and ΔT is the time to drive the NEDC for 100 km (10,720 s). $E/\Delta T$ can be interpreted as a mean power. Additionally in Table 11.4, the maximum power output of the vehicle during the NEDC and the maximum rated power are shown. $E/\Delta T$ is dramatically lower than the maximum rated power of the engine. The maximum acceleration that a vehicle can have is about two to four times (up to five to seven for sports cars) the acceleration required for the NEDC. It appears that the NEDC is rather mild in terms of acceleration and power required. Driver/customers require extremely high acceleration and power.

In [80], referring to the NEDC, the energy required to travel 100 km has been computed in case the recuperation of the kinetic energy could be obtained with perfect efficiency ($\eta = 1$) and with our increasing mass of the vehicle due to the fitting into the vehicle of the recuperation system. The formula reads

$$E_{rec} = A \cdot C_x \cdot 2.2 \times 10^4 + m \cdot f_R \cdot 9.81 \times 10^2 \text{ kJ/100 km} \tag{11.33}$$

TABLE 11.4
Mean Power Required to Complete the NEDC, Maximum Power during the NEDC and Maximum Rated Power

Vehicle Type	$E/\Delta T$ for NEDC (Mean Power) (kW)	Maximum Power during NEDC (kW)	Maximum Rated Power (kW)
Midrange	4	14	140
Compact	3	13	55
Sports	4	16	310
SUV	7	28	200

Data in Table 11.3.

Computing E_{rec}/E, it appears that, even in the ideal case, such a ratio is not lower than 80%. Considering the actual reduced efficiency of recuperation ($\eta = 0.6$) and the mass increment of the vehicle for the presence on board of the recuperation system (≈100 kg), $E_{rec}/E \approx 90\%$. In this case, only a 10% saving could be expected, that is, we cannot expect a dramatic reduction of fuel consumption from hybrid vehicles.

The aforementioned Equations 11.31 and 11.32 are fundamental for the concept design of the powertrain of vehicles with either internal combustion engine or electric or hybrid powertrain.

More information on powertrain can be found in [2,4,5,12,13,81–83].

11.3.5 Traction and Braking

Let us consider Figure 11.28. A generic two-axle vehicle with mass m and wheelbase l is shown. The center of gravity is at height h from the ground. The vehicle runs at speed v. The horizontal distances of the center of gravity from, respectively, the front and rear axles are l_F, l_R. The inertial forces are ma_x and mg.

Let us consider Figure 11.28a; the vehicle is braking, and the longitudinal deceleration is a_x. On the front axle, the forces F_{xF} and F_{zF} are acting, respectively, in longitudinal and vertical directions; correspondingly, F_{xR} F_{zR} act at the rear axle.

In the following derivations, the vectors are positive if directed as shown in Figure 11.28a.

Let us write the rotation equilibrium equation around an axis passing by point A and divide by l:

$$F_{zR} = mg\frac{l_F}{l} - ma_x\frac{h}{l}$$

When locking (at braking) or full slip (traction) is reached,

$$F_{zR} = \frac{F_{xR}}{\mu_{max}} \quad \text{and} \quad F_{zF} = \frac{F_{xF}}{\mu_{max}}$$

μ_{max} is the *maximum* friction coefficient between tire and ground; by hypothesis, it is *equal* for rear and front axle tires. Then

$$F_{zR} = mg\frac{l_F}{l} - ma_x\frac{h}{l} = \frac{F_{xR}}{\mu_{max}} \tag{11.34}$$

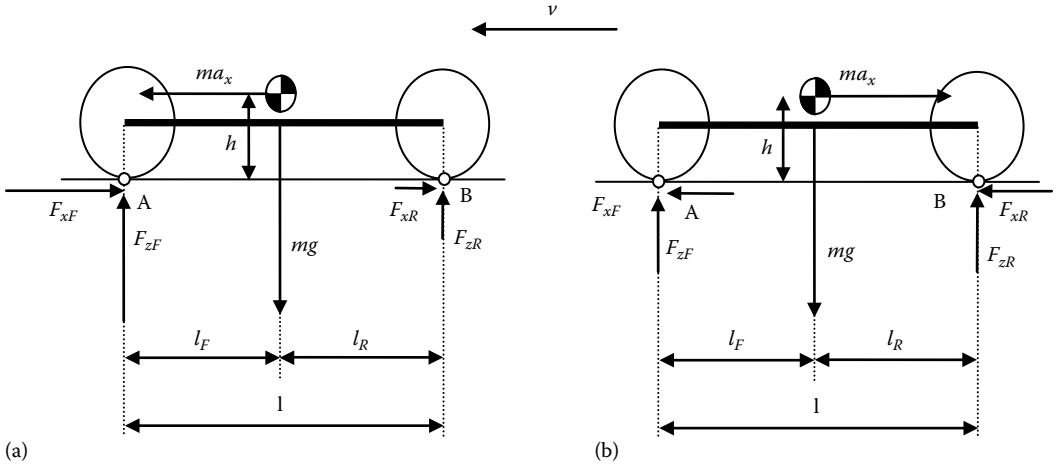


FIGURE 11.28 Forces in longitudinal and vertical direction acting on a vehicle. Aerodynamic forces are neglected. (a) Braking and (b) traction.

Repeating the aforementioned derivation with reference to point B,

$$F_{zF} = mg\frac{l_R}{l} + ma_x\frac{h}{l} = \frac{F_{xF}}{\mu_{max}} \tag{11.35}$$

Summing up Equations 11.34 and 11.35,

$$mg\frac{l_F}{l} + mg\frac{l_R}{l} = \frac{F_{xF}}{\mu_{max}} + \frac{F_{xR}}{\mu_{max}}$$

but as

$$l_F + l_R = l \quad mg = \frac{F_{xF}}{\mu_{max}} + \frac{F_{xR}}{\mu_{max}}$$

that is

$$\mu_{max} = \frac{F_{xF} + F_{xR}}{mg} \tag{11.36}$$

or $F_{xF} + F_{xR} = mg\mu_{max}$, that is, the longitudinal forces at the tires are equal to $\mu_{max}mg$.

Let us now write the equilibrium in longitudinal direction:

$$F_{xF} + F_{xR} = ma_x \tag{11.37}$$

$$a_x = \frac{F_{xF} + F_{xR}}{m} \tag{11.38}$$

Dividing Equation 11.36 by Equation 11.38, we obtain

$$\frac{\mu_{max}}{a_x} = \frac{m}{mg}$$

so

$$a_x = \mu_{max}g \tag{11.39}$$

that is, the limit longitudinal acceleration of a vehicle is $\mu_{max}g$.

Subtracting Equation 11.35 from Equation 11.34

$$mg\frac{l_F}{l} - mg\frac{l_R}{l} - 2ma_x\frac{h}{l} = \frac{F_{xR}}{\mu_{max}} - \frac{F_{xF}}{\mu_{max}}$$

eliminating μ_{max} by Equation 11.36, and rearranging

$$mg\left(\frac{l_F}{l} - \frac{l_R}{l}\right) - 2ma_x\frac{h}{l} = \frac{F_{xR} - F_{xF}}{F_{xF} + F_{xR}}mg$$

eliminating a_x by Equation 11.38

$$mg\left(\frac{l_F}{l}-\frac{l_R}{l}\right)-2\left(F_{xF}+F_{xR}\right)\frac{h}{l}=\frac{F_{xR}-F_{xF}}{F_{xR}+F_{xR}}mg$$

then (if $F_{xF}+F_{xR}\neq 0$)

$$mg\left(\frac{l_F}{l}-\frac{l_R}{l}\right)\left(F_{xF}+F_{xR}\right)-2\left(F_{xF}+F_{xR}\right)^2\frac{h}{l}=\left(F_{xF}-F_{xR}\right)mg$$

that is,

$$\begin{aligned}\left(F_{xF}+F_{xR}\right)^2&=\frac{l}{2h}\left(mg\left(\frac{l_F}{l}-\frac{l_R}{l}\right)\left(F_{xF}+F_{xR}\right)-\left(F_{xF}-F_{xR}\right)mg\right)\\&=\frac{mgl}{2h}\left(F_{xF}\frac{l_F}{l}+F_{xR}\frac{l_F}{l}-F_{xF}\frac{l_R}{l}-F_{xR}\frac{l_R}{l}-F_{xR}+F_{xF}\right)\\&=\frac{mgl}{2h}\left(F_{xF}\left(\frac{l_F}{l}-\frac{l_R}{l}+\frac{l}{l}\right)+F_{xR}\left(\frac{l_F}{l}-\frac{l_R}{l}-\frac{l}{l}\right)\right)\end{aligned}$$

As $-l_R+l=l_F$ and $l_F-l=-l_R$, then

$$\frac{mgl}{2h}\left(F_{xF}\left(\frac{2l_F}{l}\right)+F_{xR}\left(\frac{-2l_R}{l}\right)\right)\text{ and, finally,}$$

$$\left(F_{xF}+F_{xR}\right)^2=\frac{mg}{h}\left(l_FF_{xF}-l_RF_{xR}\right)\tag{11.40}$$

Equation 11.40 represents *the relationship that has to hold between longitudinal forces at, respectively, the front and rear axles when the friction at the tire–ground interface is exploited at the maximum extent* and the maximum friction coefficient is equal at the front and rear axles. Equation 11.40 has been obtained just manipulating equilibrium equations (Equations 11.34 through 11.37) and considering the limit conditions $F_{zF}=F_{xF}/\mu_{max}$, $F_{zR}=F_{xR}/\mu_{max}$. Notice that μ_{max} nor a_x appears in Equation 11.40. To plot Equation 11.40, let us write

$$F_{xR}^2+2F_{xR}F_{xF}+\frac{mg}{h}l_RF_{xR}-\frac{mg}{h}l_RF_{xF}+F_{xF}^2=0$$

F_{xR} can be obtained given F_{xF}. Figure 11.29 shows the plot of Equation 11.40.

The part of the parabola lying in the first quadrant ($F_{xF}>0$, $F_{xR}>0$) refers to braking and has been presented in Chapter 27. In Figure 11.29, not only positive (braking) forces F_{xF}, F_{xR} are plotted but also negative forces appear in the third quadrant (where $F_{xF}<0$, $F_{xR}<0$). Such forces (plotted in Figure 11.28b) are related to *traction for all-wheel-drive vehicles*. (There is no engineering interest in the case in which $F_{xF}>0$, $F_{xR}<0$ and $F_{xF}<0$, $F_{xR}>0$.) Equation 11.40 establishes the ideal torque distribution between front and rear axles to obtain the maximum longitudinal acceleration or deceleration.

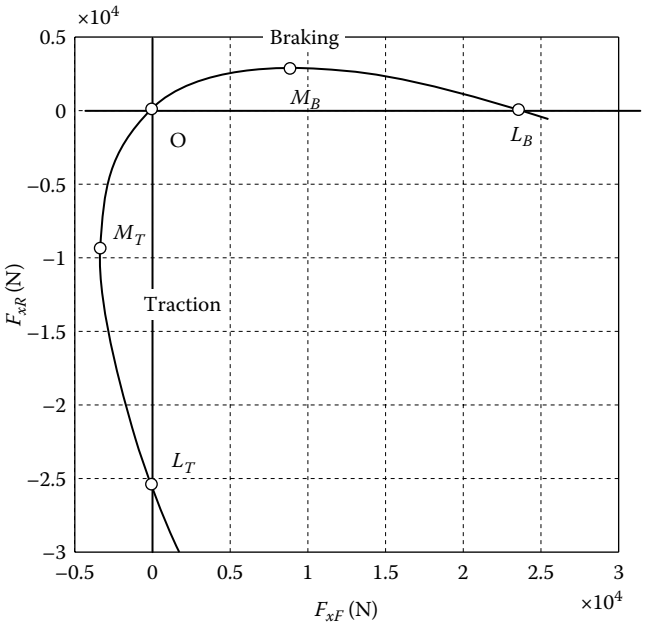


FIGURE 11.29 Plot of Equation 11.40, m = 1000 kg, l_F = 1.2 m, l_R = 1.3 m, and h = 0.5 m.

The coordinates of points O, M_B, L_B, M_T, L_T are given in Table 11.5. Such points are important to define the architecture of both braking systems and all-wheel-drive systems of cars. Usually the limit working conditions for cars are nearby points M_B and M_T; the ratios

$$\frac{F_{xFM_B}}{F_{xRM_B}} = \frac{2 \cdot l_R}{l_F} + 1, \quad \frac{F_{xFM_T}}{F_{xRM_T}} = \frac{l_R}{\left(l_R + 2l_F\right)}$$

give a guess of the maximum forces acting at front and rear axles (for cars). If $l_F = l_R$, then $F_{xFM_B}/F_{xRM_B} = 3$ and $F_{xFM_T}/F_{xRM_T} = 1/3$. The deceleration at point M_B is (Figure 11.29) $a_x = ½l_F g/h$, correspondingly, at M_T $a_x = -½l_R g/h$.

The axis of the parabola is $F_{xR} = -F_{xF} + (mg/4h)(l_F - l_R)$ (see Figure 11.30). If $l_F = l_R$, the axis takes the equation $F_{xF} = -F_{xR}$. Additionally in the origin

$$\frac{dF_{xR}}{dF_{xF}} = -1 + \frac{l}{l_R}$$

TABLE 11.5
Coordinates of Limit Points of Figure 11.29

	O	M_B	L_B	M_T	L_T
F_{xF}	0	$\frac{1}{4}\frac{mg \cdot l_F \cdot (2l_R + l_F)}{h \cdot l}$	$\frac{mg \cdot l_F}{h}$	$-\frac{1}{4}\frac{mg \cdot l_R^2}{h \cdot l}$	0
F_{xR}	0	$\frac{1}{4}\frac{mg \cdot l_F^2}{h \cdot l}$	0	$-\frac{1}{4}\frac{mg \cdot l_R \cdot (l_R + 2l_F)}{h \cdot l}$	$-\frac{mg \cdot l_R}{h}$

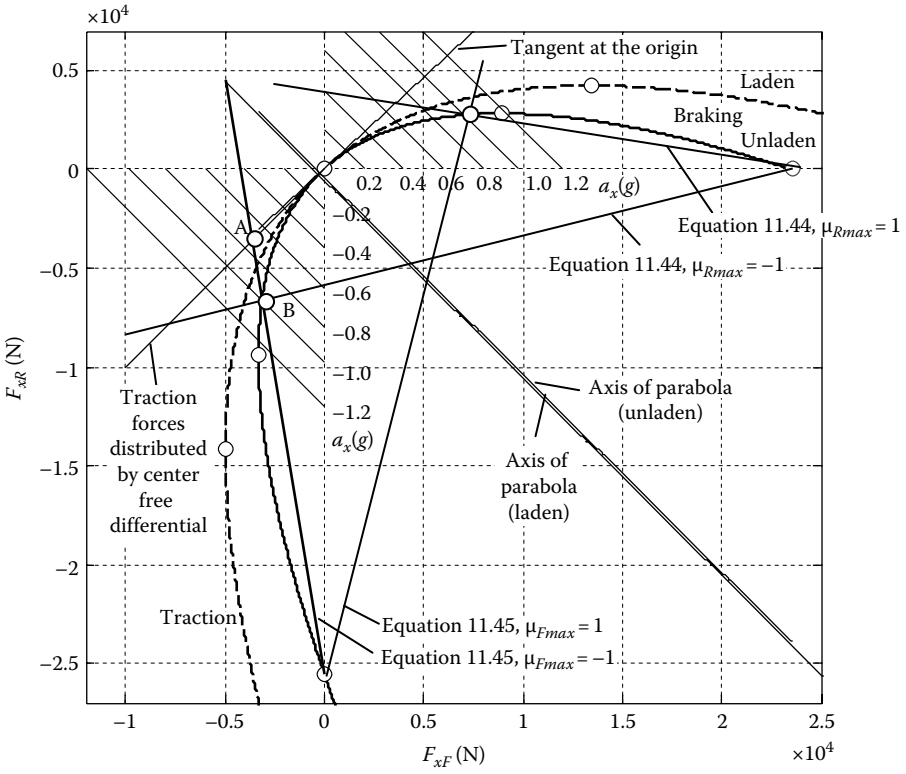


FIGURE 11.30 Traction–braking forces for optimal acceleration–deceleration performance, data in Figure 11.29 (data refer to the unladen vehicle, mass of the laden vehicle m = 1500 kg).

that is, all of the parabolas that can be obtained for different values of m have the same tangent at the origin (see Figure 11.30). If $l = 2l_R$ (or $l_F = l_R$), then $dF_{xR}/dF_{xF} = 1$.

Let us consider how Equation 11.40 is transformed when the friction coefficient is different for front and rear axles. Let us write the moment equilibrium around point A (Figure 11.28a) and divide by $lF_{zR} = mg(l_F/l) - ma_x(h/l)$. At limit condition, $F_{zR} = F_{xR}/\mu_{Rmax}$ and $F_{zF} = F_{xF}/\mu_{Fmax}$, where $\mu_{i\text{-}max}(i = F, R)$ is the maximum friction coefficient at the tire–ground interface. Let us now suppose that at the rear axle, the maximum longitudinal force is always attained, that is, $F_{zR} = F_{xR}/\mu_{Rmax}$, instead at the front axle whichever longitudinal force could be applied $F_{zF} = F_{xF}/\mu_F$ with $0 \le \mu_F < \infty$. Thus,

$$F_{zR} = mg\frac{l_F}{l} - ma_x\frac{h}{l} = \frac{F_{xR}}{\mu_{Rmax}} \tag{11.41}$$

Repeating the same derivations with reference to point B, we find

$$F_{zF} = mg\frac{l_R}{l} + ma_x\frac{h}{l} = \frac{F_{xF}}{\mu_F} \tag{11.42}$$

Summing up Equations 11.41 and 11.42, we obtain

$$mg = \frac{F_{xF}}{\mu_F} + \frac{F_{xR}}{\mu_{Rmax}}$$

that is,

$$\frac{F_{xF}}{\mu_F} = mg - \frac{F_{xR}}{\mu_{Rmax}}$$

and

$$\frac{1}{\mu_F} = \frac{1}{F_{xF}}\left(mg - \frac{F_{xR}}{\mu_{Rmax}}\right)$$

and

$$\mu_F = \frac{F_{xF}}{\left(mg - \left(F_{xR} / \mu_{Rmax}\right)\right)} = \frac{\mu_{Rmax} F_{xF}}{\mu_{Rmax} mg - F_{xR}} \tag{11.43}$$

Recalling the equilibrium equation in longitudinal direction,

$$F_{xF} + F_{xR} = ma_x \tag{11.37}$$

or

$$a_x = \frac{F_{xF} + F_{xR}}{m} \tag{11.38}$$

Subtracting Equation 11.42 from Equation 11.41, we obtain

$$mg\left(\frac{l_F}{l} - \frac{l_R}{l}\right) - 2ma_x \frac{h}{l} = \frac{F_{xR}}{\mu_{Rmax}} - \frac{F_{xF}}{\mu_F}$$

Eliminating μ_F by means of Equation 11.43 and eliminating a_x by means of Equation 11.38, we obtain

$$mg\frac{l_F}{l} - mg\frac{l_R}{l} - 2(F_{xF} + F_{xR})\frac{h}{l} = \frac{F_{xR}}{\mu_{Rmax}} - \frac{F_{xF}}{\left(\mu_{Rmax} F_{xF} / \mu_{Rmax} mg - F_{xR}\right)} = \frac{F_{xR}}{\mu_{Rmax}} - \frac{\mu_{Rmax} mg - F_{xR}}{\mu_{Rmax}}$$

Rearranging,

$$mg\frac{l_F}{l} - mg\frac{l_R}{l} - 2\frac{h}{l}F_{xF} - 2\frac{\mathrm{h}}{\mathrm{l}}F_{xR} = \frac{F_{xR}}{\mu_{Rmax}} - mg + \frac{F_{xR}}{\mu_{Rmax}}$$

and

$$-2\frac{h}{l}F_{xR} - 2\frac{F_{xR}}{\mu_{Rmax}} = -mg - mg\frac{l_F}{l} + mg\frac{l_R}{l} + 2\frac{h}{l}F_{xF}$$

and

$$\left(\frac{1}{\mu_{Rmax}}+\frac{h}{l}\right)F_{xR}=\frac{l_F}{l}mg-\frac{h}{l}F_{xF}$$

and

$$F_{xR}=\left(\frac{l_F}{l}mg-\frac{h}{l}F_{xF}\right)\left(\frac{1}{\mu_{Rmax}}+\frac{h}{l}\right)^{-1} \tag{11.44}$$

Equation 11.44 represents *the relationship that has to hold between longitudinal forces at, respectively, the front and rear axles when the friction at rear is exploited at the maximum extent.*

Let us now suppose that the tire–ground friction limit is reached at the front axle $F_{zF}=F_{xF}/\mu_{Fmax}$, instead, at the rear axle whichever longitudinal force could be applied $F_{zR}=F_{xR}/\mu_R$ with $0\le\mu_R<\infty$. Repeating the aforementioned procedure, we find

$$F_{xR}=\frac{l}{h}\left(\frac{1}{\mu_{Fmax}}-\frac{h}{l}\right)F_{xF}-\frac{l_R}{h}mg \tag{11.45}$$

Equation 11.45 represents *the relationship that has to hold between longitudinal forces at, respectively, the front and rear axles when the friction at front is exploited at the maximum extent.*

Notice that Equations 11.44 and 11.45 have been derived referring to braking. To describe traction (defined by negative forces $F_{xF}<0$, $F_{xR}<0$), in the derivation of Equations 11.44 and 11.15, the relations describing wheel slipping are

$$F_{zF}=-\frac{F_{xF}}{\mu_R},\quad F_{zR}=-\frac{F_{xR}}{\mu_R}$$

that formally imply $\mu_{Fmax}<0$, $\mu_{Rmax}<0$.

Let us derive the maximum acceleration reachable by a rear-wheel-drive car, under fundamental hypothesis that the engine torque is unlimited. From Equation 11.44, setting $F_{xF}=0$ and $F_{zR}=-F_{xR}/\mu_R$, from Equation 11.38, we find $a_x=F_{xR}/m$; combining with Equation 11.44, we find

$$a_{xRmax}=\frac{F_{xR}}{m}=\left(\frac{l_F}{l}g\right)\left(\frac{1}{-\mu_{Rmax}}+\frac{h}{l}\right)^{-1}$$

Sports cars have a very low h, with l_F that is as high as possible. Since l is proportional to the mass m, high l implies high mass, which implies a powerful (and heavy) engine.

Correspondingly, the maximum acceleration reachable by a front-wheel-drive car reads (starting from Equation 11.45)

$$a_{xFmax}=\left(\frac{l_R}{l}g\right)\left(\frac{1}{-\mu_{Fmax}}-\frac{h}{l}\right)^{-1}$$

The last two equations give hints on vehicle architecture to obtain the proper acceleration performance. If the question is comparing the maximum attainable acceleration of a front-wheel-drive car

with the maximum attainable acceleration of a rear-wheel-drive car, if $\mu_{Fmax} = \mu_{Rmax} = \mu$ and if l, h, l_F, and l_R are the same for the two vehicles, then

$$\frac{a_{xFmax}}{a_{xRmax}} = \left(\frac{l_R}{l_F}\right)\left(\frac{l - h\mu}{l + h\mu}\right) \tag{11.46}$$

Only if $l_R/l_F < (l - h\mu/l + h\mu)^{-1}$, the maximum attainable acceleration of the rear-wheel-drive car is the highest.

Figure 11.30 shows the plot of, respectively, Equations 11.37, 11.40, 11.44, and 11.45. Equation 11.40 is plotted referring both to an unladen vehicle and to a laden vehicle; Equations 11.44 and 11.45 are plotted for $\mu_{Fmax} = \mu_{Rmax} = 1$ (braking) and for $\mu_{Fmax} = \mu_{Rmax} = -1$ (traction); Equation 11.37 is plotted referring to longitudinal decelerations or accelerations ($a_x = 0.2\ldots1.2$, or $a_x = -0.2\ldots-1.2$). The line that is tangent at the origin to both of the two curves "laden" and "unladen" is shown, and its slope is almost equal to 1 (as $l \approx 2l_R$). The respective axes of the two parabolas "laden" and "unladen" are shown, and they are very near.

Let us consider Figure 11.30. Let us imagine that $|\mu_{Fmax}| = 1$ and $|\mu_{Rmax}| = 1$. In traction condition for an all-wheel-drive car, point B represents the maximum possible acceleration (provided that the engine power is sufficient; actually the acceleration would be equal to 1 g, a performance hardly achievable even by a very efficient sports car). In this case, both the front and the rear axles come together at a full slip. Let us imagine that the architecture of the all-wheel-drive car is conceived with a central free differential. In this case, the longitudinal force distribution is equal on the front and rear axles, and such a condition is represented by the line equally sloped shown in the third quadrant of the plot in Figure 11.30. Increasing the module of the forces at the axles, point A is reached from the origin; at point A, the slipping of front axle is reached, and the rear axle is still not slipping. Since the maximum longitudinal force at the front axle is reached, the free differential prevents to increase the force at the rear axle even if an additional torque would be applied by the engine (such an additional torque input would imply unsteady motion conditions of the driveline that is not taken into consideration in our model). The maximum acceleration attainable would be about 0.7 g instead of 1 g. Figure 11.30 shows that the distribution of longitudinal forces at the axles during traction to obtain the best performance in terms of maximum acceleration is not the distribution provided by a free differential connecting the front and the rear axle. Actually, the best distribution should be defined by Equation 11.40.

This occurrence has important implications in the conceptual design of the architecture of all-wheel-drive road vehicles. Actually, at least a fixed unequal distribution of longitudinal forces at the front and rear axles is necessary (about 40% at front 60% at rear), as it is often adopted in very powerful cars (for more details on driveline architecture, see [2,85]).

Obviously, the maximum longitudinal acceleration depends on the friction available. In case friction at the tire–ground interface were extremely high, the maximum longitudinal acceleration would be obtained with rear-wheel drive only (i.e., $F_{xF} = 0$).

11.4 ACTIVE SAFETY

11.4.1 Handling

In Chapter 27, the main concepts referring to handling have been given. Here, the influence on handling of some parameters related to vehicle architecture will be highlighted.

Handling of road vehicles (running on even road) depends mostly on

- Tire characteristics
- Inertial and geometric data (mass, center of gravity location, wheelbase)
- Drive axle (front- or rear- or all-wheel drive) and torque vectoring
- Suspension kinematics and elastokinematics

A simplified but extremely powerful way to take into account all of the relevant parameters or nonlinearities influencing the horizontal motion of vehicles is provided by the *handling diagram* theory. Such a theory describes the handling and stability properties of a vehicle running on an even surface. The handling diagram theory is briefly reported in Chapter 27, Section 27.3.3; full details are given in [8]. In the following sections, the handling diagram theory will be used to highlight the influence of parameters related to vehicle architecture on active safety. For further reading on active safety, see [1,2,7,11,86].

11.4.1.1 *Tire Characteristics*

Surely, the nonlinear tire cornering characteristics play *the* fundamental role referring to handling. Let us consider Figure 11.31a; the forces acting on an actual car (left) and the corresponding resultants acting at the axles (right) are shown. Such a correspondence suggests the introduction of the simple two-wheel model depicted in Figure 11.31b (Chapter 27) for studying the handling behavior of road vehicles. Let us consider the well-known tire cornering characteristic (briefly, tire characteristic) at pure lateral slip, that is, let us consider the relation between lateral force and lateral slip $F_y = F_y(\alpha)$. Let us consider the two-wheel model depicted in Figure 11.31b; for the model to be consistent with reality, the resultant lateral force at the rear axle is to be computed as the sum of the lateral forces at the two rear tires of the actual vehicle (e.g., considering the rear axle, $F_{yR} = 2F_y(\alpha)$). Approximately the same holds for the front axle (precisely, the resultant lateral force at the front axle can be computed as the sum of the lateral forces at the two front tires if a vanishing longitudinal force is acting at the front tires and—as it usually occurs—δ_F is small; see Chapter 27). The relationship between axle cornering characteristic (defined as the resultant lateral forces acting at the axle as function of lateral slip at the axle center) and tire characteristic is obvious for the two-wheel model in Figure 11.31b (e.g., considering the rear axle, $F_{yR} = 2F_y(\alpha)$) in linear form $F_{yR} = 2C_{\alpha R}\alpha$ ($C_{\alpha R}$ is the rear tire cornering stiffness; for definition of the cornering stiffness, see Chapters 17 and [8]). The relationship between axle cornering characteristic (briefly, axle characteristic) and tire characteristic is less obvious if the actual vehicle (Figure 11.31a) or a more complex vehicle model is considered. In this case, many different vehicle parameters, such as the axle kinematic or elastokinematic, can have a substantial influence in defining the said relationship (the more relevant among such influences will be dealt with in the subsequent sections). It is anyway a matter of fact that such influences do not modify substantially the original relation between axle characteristic and tire characteristic. Thus, the resultants of lateral forces at the axles are strictly related to tire characteristics.

In the handling diagram theory (see Chapter 27, Section 27.3.3 [8]), reference is made to the so-called effective axle characteristics that are the previously introduced axle characteristic divided by the resultant of the vertical forces at the same axle F_{yi}/F_{zi} ($i = F, R$) (e.g., referring to rear axle and to Figure 11.32, $F_{yR}/F_{zR} = F_{yR}(\alpha)/F_{zR}$, in linear form $F_{yR}/F_{zR} = C_R\alpha/F_{zR} = 2C_{\alpha R}\alpha/F_{zR}$). Referring to Figure 11.31c and considering the effective axle characteristics F_{yF}/F_z, F_{yR}/F_z, (subscripts F = front, R = rear), it is clear how the front and rear effective axle characteristics influence the handling diagram [8]. Actually, by inspection of Figure 11.31c, one can easily see that the curve on the left (the so-called handling curve) is obtained *by subtracting horizontally the two curves on the right (i.e., the effective axle characteristics)*, at a given value of F_{yi}/F_{zi}. Since

$$\frac{F_{yF}}{F_{zF}} = \frac{F_{yR}}{F_{zR}} = \frac{a_y}{g} = \frac{v_x^2}{Rg} \quad \text{and} \quad \alpha_F - \alpha_R = \delta_F - \frac{l}{R} = \delta_F - \delta_{F_0}$$

the handling curve of Figure 11.31c establishes a relationship between the lateral acceleration a_y/g (expressed in nondimensional form as a_y is divided by the ground acceleration g) and the steering angle δ_F. (For the sake of simplicity, the aforementioned formula has been derived neglecting the aerodynamic forces and the rolling resistances—such effects can be introduced with a proper reasoning that is dealt with in Section 11.4.1.3).

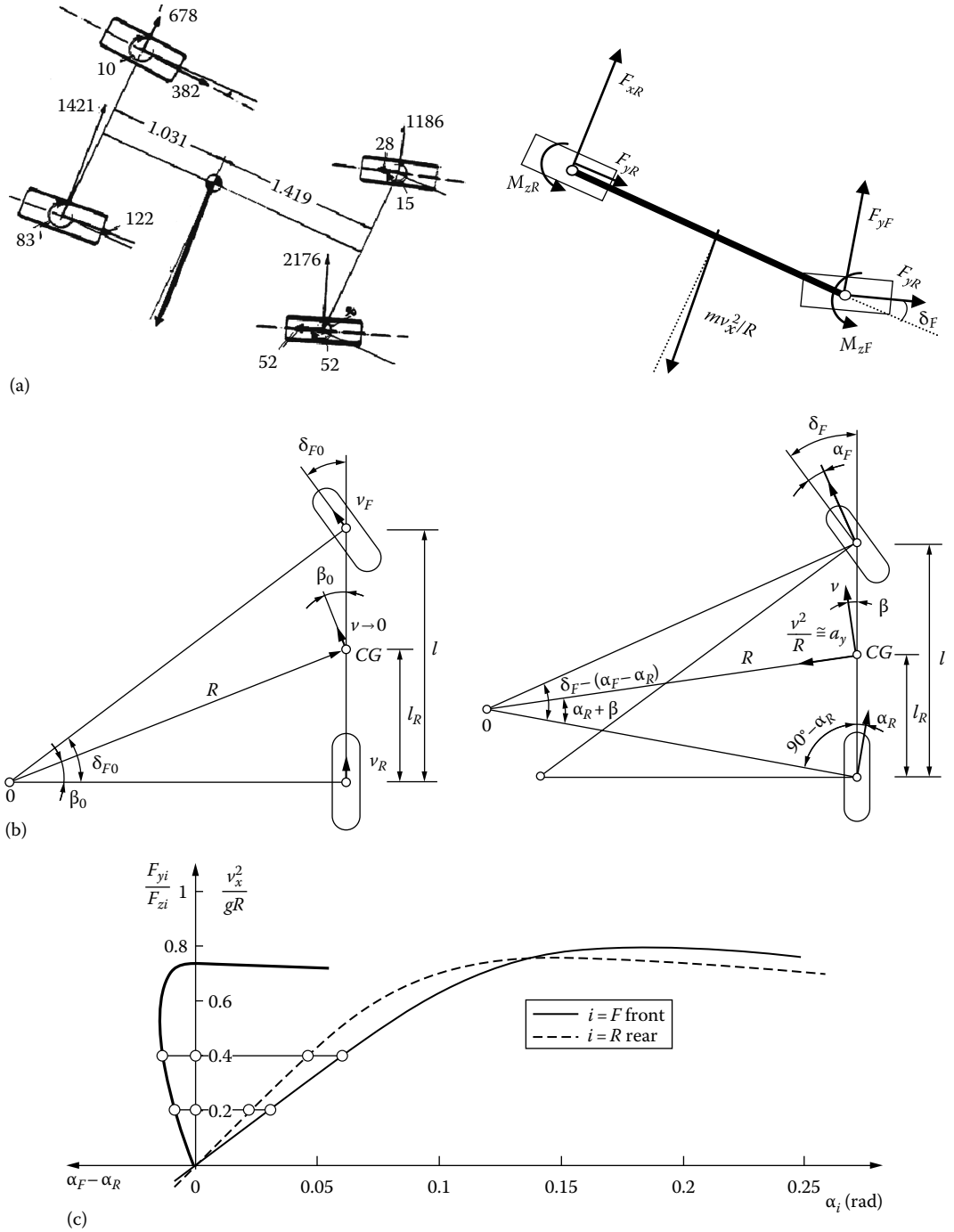


FIGURE 11.31 (a) Forces acting on an actual car (left) and corresponding resultants acting at the axles (right); aerodynamic forces are neglected. (b) Simple car model (see Chapter 27). (c) Front and rear tire characteristics and corresponding handling diagram. (Adapted from Figure 27.30.) The left curve is obtained by subtracting the effective front axle characteristic from the rear one at a given level of lateral acceleration (see, e.g., the segments at 0.2 and 0.4).

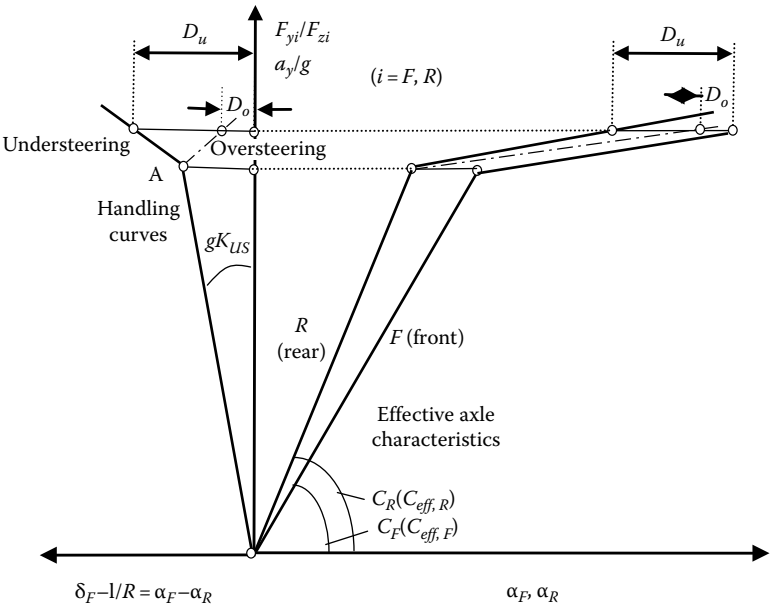


FIGURE 11.32 Substantial variation of the handling curve due to slight variation of the rear effective axle characteristic at high lateral acceleration. F_{yi} resultant of lateral forces at ith axle, F_{zi} resultant of vertical forces at ith axle, a_y centripetal acceleration (g gravity), l wheelbase, R cornering radius, and δ_F front-wheel steer. The effective axle characteristics have been depicted in simplified multi-linear form. K_{US} is the understeering gradient (see Chapter 27 and Equation 11.48).

The front and rear effective axle characteristics can be very much different from the ones depicted in Figure 11.31c. In [8], a number of different combinations of tire characteristics are presented together with the corresponding handling diagrams.

Any parameter (of the vehicle or of the tire) that influences the tire characteristic does influence the axle effective characteristic and consequently the handling diagram, that is the dynamic behavior of the vehicle.

Considering the so-called tire brush model [8], the cornering stiffness C_α can be approximated as

$$C_\alpha = 2 \cdot c_{py} L^2 = 2 \cdot c_{pby} b_c L^2$$

where

c_{py} is the lateral stiffness of the tire tread per unit length
c_{pby} is the lateral stiffness of the tire tread per unit area (the area refers to the contact patch area)
L is the half length of the contact area (see Figure 11.22)
b_c is the tire tread cross width (Figure 11.19)

Higher cross width b_c is related to higher cornering stiffness C_α. When a tire has to be chosen to be fitted into a car, usually, the driving condition is that the tire has to carry the vertical load $F_{z,stat} \cong 2 p_i b_c \sqrt{2 R w_{max}} \cong 2 p_i b_c L$ (Equation 11.30); the tire manufacturers provide different combinations of parameters p_i, b_c, and R, to carry the same vertical force $F_{z,stat}$. Among such a variety of solutions, the designer can choose the tire that best fits the needs, considering not only the cornering stiffness but also the rolling resistance (which is—among other factors—proportional to $b_c t$; see Figure 11.19 and Section 11.3.3.2) and the radial stiffness (which is approximately proportional to $p_i \cdot b_c \cdot \sqrt{R}$; see Section 11.3.3.1).

It is now clear how the dimensions of a tire can influence the tire characteristic, that is, the effective axle characteristic, and finally the handling diagram. In the subsequent sections, the main parameters related to vehicle architecture that influence axle effective characteristics will be taken into account.

Let us consider Figure 11.32; the handling diagram theory immediately highlights that the vehicle handling (and stability) is very much influenced by slight variations of tire characteristics, especially at high lateral acceleration. Let us depict the front and rear effective axle characteristics by means of straight segments; Figure 11.32 represents a simplified version of both the axle characteristics and the handling diagram that are shown in Figure 11.31c. Figure 11.32 shows that a slight variation of rear axle characteristic causes, after point A, the vehicle to change from understeering to oversteering (for the definition of understeering and oversteering, see Chapter 27). This means that at high lateral accelerations, a slight variation of the axle characteristics can cause a huge change of the overall vehicle dynamic behavior. Since the effective axle characteristics are influenced by chassis parameters (e.g., suspension kinematics influence the tire forces that influence effective axle characteristic), at high lateral accelerations, slight variation of even one of those parameters may influence considerably the stability of the vehicle (see Section 11.4.1.4).

11.4.1.2 Inertial and Geometric Data

To assess the handling performance of a car (or more generally of a ground road vehicle), the *understeering gradient* K_{US} (see Figure 11.31) is fundamental

$$K_{US} = \frac{m}{l}\frac{C_R l_R - C_F l_F}{C_R C_F} \tag{11.47}$$

where

m is the vehicle mass

l is the wheelbase

C_i and l_i ($i = F, R$ F = front, R = rear) are, respectively, the effective axle characteristic and the horizontal distance of the center of gravity from the ith axle center

The understeering gradient defines the handling curve at relatively low levels of centripetal acceleration (Figure 11.32); thus, the study of the understeering gradient is directly related to the study of the handling diagram.

Additionally, to assess the active safety, the following indices can be considered (see Chapter 27):
Yaw velocity gain (yaw velocity at steady state per unit of front-wheel steer angle):

$$G_{r,st} = \frac{v_x}{l + K_{US} v_x^2} \tag{11.48}$$

where v_x is the forward vehicle speed.

Damping ratio of *both* the yaw rate r (Figure 11.39) and the sideslip angle β (Figure 11.31b) responses to a steering input:

$$D = \frac{K_1}{\sqrt{K_2}} \tag{11.49}$$

where K_1 and K_2 are taken from the differential equations governing the yaw rate r and the sideslip angle β caused by a steering input

$$\ddot{\beta} + 2K_1\dot{\beta} + K_2\beta = \frac{C_F}{mv_x}\dot{\delta}_F - \frac{C_F\left(l_F mv_x^2 - C_R l_R l\right)}{I_Z mv_x^2}\delta_F$$

$$\ddot{r} + 2K_1\dot{r} + K_2 r = \frac{l_F C_F}{I_Z}\dot{\delta}_F + \frac{C_F C_R l}{I_Z mv_x}\delta_F$$

$$K_1 = \frac{I_Z\left(C_F + C_R\right) + m\left(C_F l_F^2 + C_R l_R^2\right)}{2I_Z mv_x}$$

$$K_2 = \frac{l^2 C_F C_R + \left(C_R l_R - C_F l_F\right)mv_x^2}{I_Z mv_x^2}$$

See Chapter 27 (Equations 27.75 through 27.78) for the meaning of the symbols; the damping ratio by definition is $D = 2K_1 / 2 \cdot 1 \cdot \sqrt{K_2/1}$. The damping ratio gives important information to assess if a vehicle oscillates or not after a disturbance.

Let us consider the influence of the mass m, the wheelbase l, and the center of gravity location $l_F = k_l l$, ($l_R = l - l_F$) (Figure 11.31b), respectively, on the understeering gradient K_{US}, on the yaw rate gain $G_{r,st}$, and on the damping ratio D.

Such a handling assessment can be made with two different perspectives. The first perspective is the one of the designer who performs the *conceptual design* of a vehicle, that is, the perspective of the engineer who has to define the basic architecture of a fully new vehicle. The second perspective is the one of an engineer who has to perform a *robustness assessment*. In the conceptual design phase, typically, the wheelbase, the location of the center of gravity, and the mass have to be decided. In the robustness assessment, given a vehicle, the effects on handling of the loading and related displacement of the center of gravity have to be minimized.

In the following derivations, we will take into account the fact that the mass m and the wheelbase l are inherently related; a linear interpolation based on actual data (see Figure 11.42) gives $m = m(l) = k_{ml}l + q$. The convenient relation between the wheelbase and location of the center of gravity will be introduced by the two equations $l_F = k_l l$, $l_R = l - l_F$ (l_F is the horizontal distance between the center of gravity and the fore axle center; see Figure 11.31b). For a given vehicle that is to be loaded, the location of the center of gravity appears to be related to the mass m; a linear interpolation based on actual data can be used, which gives $l_F = l_F(m) = k_l l$, $k_l = k_m(m - m_0) + k_{l0}$ (see Equation 11.66 for the meaning of symbols). The relation between the moment of inertia around the vertical axis I_z and the mass is $I_z = m \cdot l_R \cdot l_F$ (see Section 11.5 where $J_{zz} = I_z$).

In order to introduce in a simple and effective way the influence on handling of the tires, let us consider two cases referring to the cornering stiffness:

$$1.\ C_F = C_F(F_z) = K_F F_{zF}, \quad C_R = C_R(F_z) = K_R F_{zR} \tag{11.50}$$

$$2.\ C_F = C_F(F_z) = K_{FS} = \text{constant}, \quad C_R = C_R(F_z) = K_{RS} = \text{constant} \tag{11.51}$$

The first case and the second case refer, respectively, to lines (1) and (2) in Figure 11.33. The first case refers to a case in which the tire is not exploited very much; on the contrary, the second case refers to a substantial exploitation of the tire. Roughly, the first case refers to relatively large tires (high cross width) and relatively small (thin) tires work in the second case.

Usually, for cars, K_i ($i = F, R$) can be found in the range $5 < K_i < 25$ (frequently, for car tires, K_i values can be found in the range 10–20; for truck tires, see [8]). K_i ($i = FS, RS$), for cars, can be found in the range 50,000 N/rad $< K_i <$ 200,000 N/rad. Commonly $K_R > K_F$ and $K_{RS} > K_{FS}$.

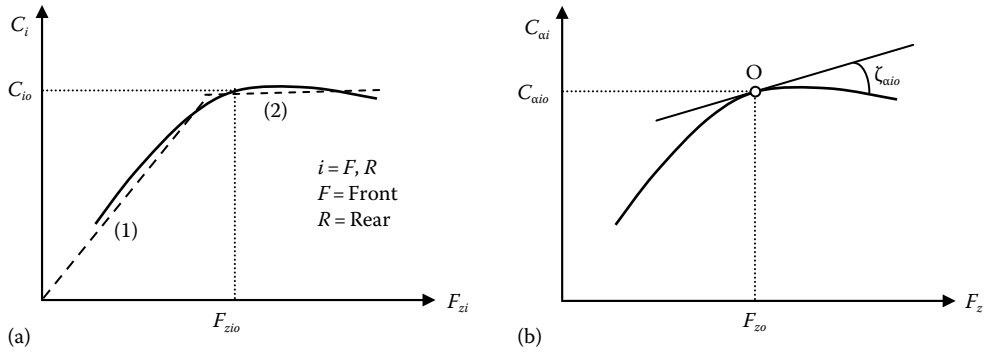


FIGURE 11.33 (a) Approximation of the effective axle characteristic C_i ($i = F,R$) (function of the vertical load at the axle F_{zi} ($i = F,R$)) by means of a multi-linear function composed by line (1) and line (2). (b) Approximation of the tire cornering stiffness by means of linearization about point O.

Referring to the two-wheel model described in Figures 11.31 and 11.35b, the vertical forces at the tires read $F_{zF} = mgl_R/l$, $F_{zR} = mgl_F/l$; so for case (1), we can write $C_F = C_F(F_z) = K_F F_{zF} = K_F(mgl_R/l)$, $C_R = C_R(F_z) = K_R F_{zR} = K_R(mgl_F/l)$.

K_{US} as a function of l, l_F, l_R, and m

Referring to Equations 11.50 and 11.51, we obtain for K_{US}

Case (1)

$$K_{US} = \frac{m}{l}\frac{C_R l_R - C_F l_F}{C_R C_F} = \frac{m}{l}\frac{K_R\left(mgl_F/l\right)l_R - K_F\left(mgl_R/l\right)l_F}{K_R\left(mgl_F/l\right)K_F\left(mgl_R/l\right)} = \frac{1}{g}\frac{K_R - K_F}{K_F K_R} = \frac{1}{gK_F} - \frac{1}{gK_R} \tag{11.52}$$

Case (2)

$$K_{US} = \frac{m}{l}\frac{C_R l_R - C_F l_F}{C_R C_F} = \frac{m}{l}\frac{K_{RS} l_R - K_{FS} l_F}{K_{RS} K_{FS}} = m\frac{K_{RS}\left(1-k_l\right) - K_{FS}k_l}{K_{RS}K_{FS}} = m\frac{K_{RS} - \left(K_{RS} + K_{FS}\right)k_l}{K_{RS}K_{FS}} \tag{11.53}$$

In case (1) (tires are not exploited very much), the understeering gradient *does not depend* on the vehicle parameters (wheelbase, location of the center of gravity, mass) but on tire characteristics only! If $K_F = K_R$, then $K_{US} = 0$.

In case (2) (tires are substantially exploited), the understeering gradient does depend on the vehicle mass and location of the center of gravity and does not depend explicitly on the wheelbase.

From a conceptual design perspective, if the mass is increased (without changing the position of the center of gravity), the understeering (or oversteering behavior) is increased. If the position of the center of gravity is moved toward the rear axle (i.e., k_l increases without increasing the mass), the understeering gradient K_{US} is decreased (oversteering is approached or made worse).

Since mass is linearly related to wheelbase ($m = m(l) = k_{ml}l + q$; see Figure 11.42), K_{US} turns out to increase linearly with the wheelbase, even if the wheelbase does not appear explicitly in Equation 11.53.

From a robustness assessment perspective, loading a given car or a van (except for some sports cars with rear engine) implies that the center of gravity moves toward the rear axle. Considering

Equation 11.66 that gives $l_F = l_F(m) = k_l l$, we can set $k_l = k_m(m - m_0) + k_{l0} = k_m(\Delta m) + k_{l0}$, and substituting into Equation 11.53,

$$K_{US} = m\frac{K_{RS}\left(1-k_m\left(m-m_0\right)-k_{l0}\right)-K_{FS}\left(k_m\left(m-m_0\right)+k_{l0}\right)}{K_{RS}K_{FS}}$$

$$= -k_m\left(\frac{1}{K_{FS}}+\frac{1}{K_{RS}}\right)\left(\Delta m\right)^2+\left(\frac{1-m_0k_m-k_{l0}}{K_{FS}}-\frac{m_0k_m+k_{l0}}{K_{RS}}\right)\Delta m+\left(\frac{m_0\left(1-k_{l0}\right)}{K_{FS}}-\frac{m_0k_{l0}}{K_{RS}}\right)$$

$$= A\left(\Delta m\right)^2+B\Delta m+C \tag{11.54}$$

where

m_0 represents the mass of the unloaded vehicle

$\Delta m = m - m_0$ represents the mass increase when the vehicle is loaded

Equation 11.54 represents a parabola pointing downward and whose vertex abscissa is at

$$\left(\Delta m\right)_v = \frac{-B}{A} = -\frac{K_{RS}\left(m_0k_m+k_{l0}-1\right)+K_{FS}\left(m_0k_m+k_{l0}\right)}{k_m\left(K_{FS}+K_{RS}\right)}$$

Since usually $(\Delta m)_v < 0$ and $C = (m_0(1 - k_{l0})/K_{FS}) - (m_0k_{l0}/K_{RS}) > 0$, K_{US} decreases if Δm increases (i.e., K_{US} decreases if the vehicle is loaded).

$G_{r,st}$ as a function of l, l_F, l_R, and m

Referring to Equations 11.50 and 11.51, we obtain for $G_{r,st}$

Case (1)

$$G_{r,st} = \frac{v_x}{l+v_x^2\left(\frac{1}{gK_F}-\frac{1}{gK_R}\right)} \tag{11.55}$$

Case (2)

$$G_{r,st} = \frac{v_x}{l+v_x^2\left(m\frac{K_{RS}-(K_{RS}+K_{FS})k_l}{K_{RS}K_{FS}}\right)} \tag{11.56}$$

The yaw rate gain $G_{r,st}$ is directly related to the understeering gradient K_{US} (see Equation 11.48), so similar comments can be made here with respect to the ones made earlier.

In case (1), the yaw rate gain does not depend on the location of the center of gravity. The yaw rate gain decreases substantially if the wheelbase increases.

From a conceptual design perspective, increasing the wheelbase implies a linear increase of the mass (see Figure 11.42), so the yaw rate gain decreases substantially if the mass increases, even if the mass does not appear explicitly in Equation 11.55.

From a robustness assessment perspective, the sensitivity to loading is null.

In case (2), the yaw rate gain does depend on the vehicle mass, location of the center of gravity, and wheelbase.

From a conceptual design perspective, if the mass is increased (without changing the wheelbase and position of the center of gravity), the yaw rate gain is decreased. If the position of the center of gravity is moved toward the rear axle (i.e., k_l increases), without increasing the mass, the yaw rate gain is increased.

From a robustness assessment perspective, given a car or a van, inserting Equations 11.54 and 11.66 into Equation 11.56 and rearranging,

$$G_{r,st} = \frac{v_x}{l + v_x^2\left(-k_m\left(\frac{1}{K_{FS}} + \frac{1}{K_{RS}}\right)(\Delta m)^2 + \left(\frac{1 - m_0 k_m - k_{l0}}{K_{FS}} - \frac{m_0 k_m + k_{l0}}{K_{RS}}\right)\Delta m + \left(\frac{m_0(1-k_{l0})}{K_{FS}} - \frac{m_0 k_{l0}}{K_{RS}}\right)\right)} \tag{11.57}$$

By inspection of Equation 11.57 and recalling the discussion about Equation 11.54, one may notice that $G_{r,st}$ increases if Δm increases.

D as a function of l, l_F, l_R, and *m*

Referring to Equations 11.50 and 11.51, we obtain for *D*

Case (1)

$$D = \frac{1}{2}\frac{g\left(K_F + K_R\right)l}{\sqrt{g\left(g \cdot l \cdot K_F K_R + v_x^2 K_R - v_x^2 K_F\right)l}} \tag{11.58}$$

Case (2)

$$\begin{aligned} D &= \frac{1}{2}\frac{\left(K_{FS} l_F + K_{RS} l_R\right)l}{\sqrt{-\left(-l^2 K_{FS} K_{RS} - m v_x^2 K_{RS} l_R + v_x^2 K_{FS} l_F\right) l_F l_R}} \\ &= \frac{1}{2}\frac{\left[\left(K_{FS} - K_{RS}\right)k_l + K_{RS}\right]l}{\sqrt{l\left(-l K_{FS} K_{RS} - m v_x^2 K_{RS} + m v_x^2 K_{RS} k_l + m v_x^2 K_{FS} k_l\right)k_l\left(-1 + k_l\right)}} \end{aligned} \tag{11.59}$$

In case (1), the damping ratio *D* does not depend on the location of the center of gravity and does not depend explicitly on the mass. *D* depends on the wheelbase, but quantitatively the effect is usually small. If $K_F = K_R$, then $D = 1$.

From a conceptual design perspective, the wheelbase and the mass increase accordingly in a linear way, so *D* increases (slightly) with the mass.

From a robustness assessment perspective, the sensitivity to loading is null.

In case (2), the damping ratio *D* depends explicitly on the wheelbase, on the location of the center of gravity, and on the mass.

From a conceptual design perspective, Figure 11.34 shows the influence of the wheelbase on the damping ratio *D* for two different vehicle speeds and two different positions of the center of gravity.

From a robustness assessment perspective, given a vehicle, inserting Equation 11.66 in Equation 11.59 leads to an involved symbolic expression that reads

$$D = \frac{f\left(\Delta m\right)}{\sqrt{g\left(\left(\Delta m\right)^4\right)}} \tag{11.60}$$

where *f* and *g* are two functions. In this case, discussing $D(\Delta m)$ is very difficult and a numerical solution is to be envisaged. Often, increasing the mass (increasing Δm), *D* decreases.

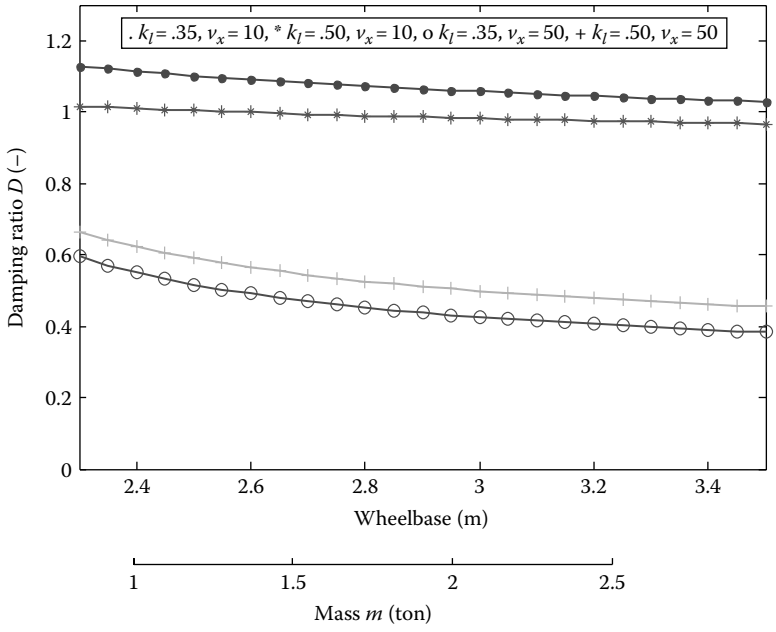


FIGURE 11.34 Plot of Equation 11.59; the damping ratio D depends mostly on the speed v_x ($[v_x]$ = m/s), less on the wheelbase (or mass) and less on the position of the center of gravity (defined by k_l). Understeering vehicle, K_{FS} = 60,000 N/rad, K_{RS} = 110,000 N/rad.

The aforementioned analyses state that working in the part (1) of tire cornering stiffness characteristic (Figure 11.33) allows to conceive a vehicle that is very much less sensitive to mass variation (and related center of gravity migration). On the contrary, a vehicle conceived to work in the part (2) of tire cornering stiffness characteristic (Figure 11.33) is sensitive to all of the vehicle geometric and inertial parameters.

11.4.1.3 Drive Axle and Torque Vectoring

Let us recall Chapter 27, Section 27.3.3, where the handling diagram theory was introduced together with the hypotheses on which such a theory is based on.

Let us consider the simple two-wheel model in Figures 11.31b and 11.35. Let us consider the longitudinal axle of such a vehicle model and the external resultant forces acting on it

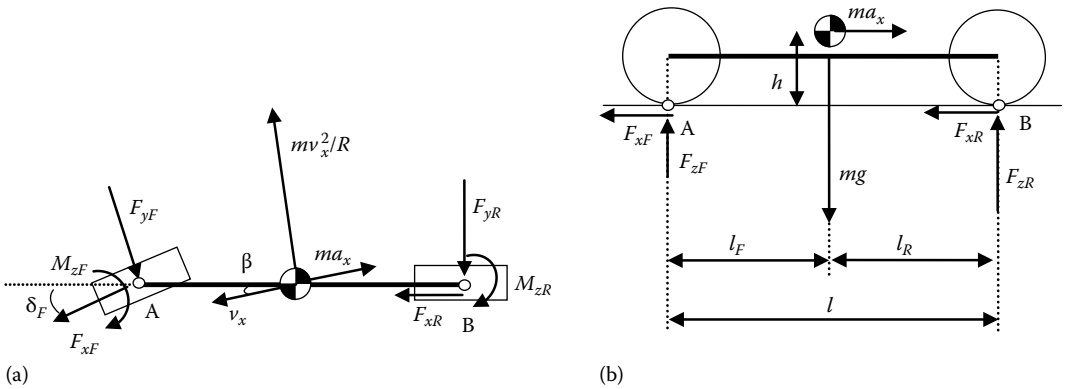


FIGURE 11.35 External resultant forces acting on a two-axle vehicle (aerodynamic effects are neglected). (a) Top view and (b) lateral view.

(aerodynamic effects are neglected). The moments M_{zF} and M_{zR} in Figure 11.35 are introduced as they represent unequal longitudinal forces at the two wheels of the same axle, thus representing a torque vectoring effect.

Let us assume that the steering angle δ_F is small (this assumption is consistent with the hypothesis given in Chapter 27, where the force $F_{xF} \sin \delta_F$ is neglected) and that the β angle is small. The rotation equilibrium around a vertical axis passing by point B (taking into account the forces and moments acting in the horizontal plane) is written in the first equation (11.61). The rotation equilibrium around a vertical axis passing by point A (taking into account the forces and moments acting in the horizontal plane) is written in the second equation (11.61). The rotation equilibrium around a horizontal axis passing by point B (taking into account the forces in the vertical plane) is written in the third equation (11.61). The rotation equilibrium around a horizontal axis passing by point A (taking into account the forces in the vertical plane) is written in the fourth equation (11.61):

$$\begin{aligned} F_{yF} l &= m \frac{v_x^2}{R} l_R + \left(M_{zF} + M_{zR}\right) \\ F_{yR} l &= m \frac{v_x^2}{R} l_F - \left(M_{zF} + M_{zR}\right) \\ F_{zF} l &+ m a_x h = m g l_R \\ \boldsymbol{F}_{zR} l &- m a_x h = m g l_F \end{aligned} \tag{11.61}$$

The first equation (11.62) is obtained by dividing the first equation (11.61) by the third equation (11.61). The second equation (11.62) is obtained by dividing the second equation (11.61) by the fourth equation (11.62). By this way, a mathematical manipulation of the equilibrium equations is performed; thus, the general validity of new relations is not corrupted:

$$\begin{aligned} \frac{F_{yF}}{F_{zF} + m a_x h / l} &= \frac{m v_x^2 / R l_R}{m g l_R} + \frac{\left(M_{zF} + M_{zR}\right)}{m g l_R} \\ \frac{F_{yR}}{F_{zR} - m a_x h / l} &= \frac{m v_x^2 / R l_F}{m g l_F} - \frac{\left(M_{zF} + M_{zF}\right)}{m g l_F} \end{aligned} \tag{11.62}$$

The aforementioned equations can be rearranged considering that

$$F_{zF} + \frac{m a_x h}{l} = F_{zF,st}$$

$$F_{zR} - \frac{m a_x h}{l} = F_{zR,st}$$

where $F_{zF,st}$ and $F_{zR,st}$ refer, respectively, to the static vertical forces at the front and rear axles. So we obtain

$$\begin{aligned} \frac{F_{yF}}{F_{zF,st}} - \frac{\left(M_{zF} + M_{zR}\right)}{m g l_R} &= \frac{v_x^2}{R g} \\ \frac{F_{yR}}{F_{zR,st}} + \frac{\left(M_{zF} + M_{zR}\right)}{m g l_F} &= \frac{v_x^2}{R g} \end{aligned} \tag{11.63}$$

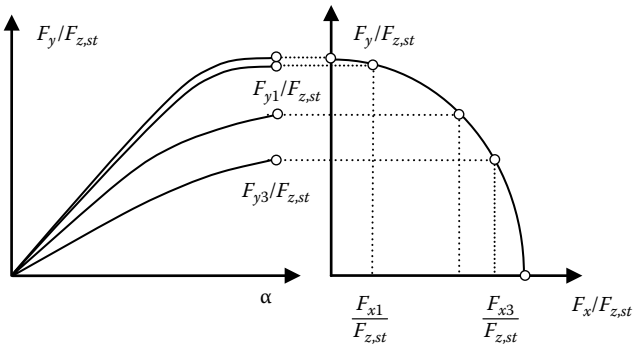


FIGURE 11.36 Left plots lateral forces at different longitudinal forces. Right plots relationship between maximum forces acting, respectively, in longitudinal (x) and in lateral direction (y) on a tire. α is the lateral slip and the suffix "*st*" means that the vertical force at the tire is constant (steady state).

Equation 11.63 corresponds to Equation 27.122 in Chapter 27 if M_{zF} and M_{zR} are neglected. Moreover, such equations have been obtained considering a motion with constant acceleration along the longitudinal axis!

This derivation is very important because it means that the handling diagram theory can give an (approximated) insight of the nonlinear dynamic behavior of vehicles accelerating into a bend. This occurrence will be exploited here to discuss the conceptual design of two-axle vehicles with different driveline architectures (front-, rear-, or all-wheel drive).

Often in the engineering practice, the handling diagrams (relation between $\delta_F - \delta_{F0}$ and centripetal acceleration) are obtained at different longitudinal accelerations.

Let us consider Figure 11.36 where a simplified graphical representation of the relationship between the longitudinal and lateral forces at a tire is given [2] (for the correct relationship, see [8]). Increasing the longitudinal force at the tire from $F_{x1}/F_{z,st}$ to $F_{x3}/F_{z,st}$ leads to the lowering of the tire characteristic $F_y/F_{z,st}(\alpha)$; actually $F_{y1}/F_{z,st}(\alpha)$ is higher than $F_{y3}/F_{z,st}(\alpha)$.

This general conclusion can be used to understand the effects on handling of the architecture of the driveline.

In Figure 11.37a, the effects on handling of front-wheel drive are highlighted. Since the longitudinal force at tires causes a lowering of the tire characteristic (Figure 11.36), the vehicle with front-wheel drive is more understeering (increased understeering gradient K_{US}) than the vehicle with vanishing forces at the front tires. Additionally, the attainable maximum centripetal acceleration is lowered with respect to the case of vanishing forces at the front tires.

In Figure 11.37b, the effects on handling of rear-wheel drive are highlighted. If the longitudinal forces at rear tires are not very high, the vehicle is less understeering (decreased understeering gradient K_{US}) than the vehicle with vanishing forces at the rear tires. If the longitudinal forces at rear tires are very high, the vehicle is oversteering (understeering gradient $K_{US} < 0$). Always the attainable maximum centripetal acceleration is lowered with respect to the case of vanishing forces at the rear tires.

In Figure 11.37c, the effects on handling of all-wheel drive are highlighted. Usually the torque distribution between the front and rear axles is chosen to keep the vehicle almost insensitive to the application of longitudinal forces at the front and rear tires. Always the attainable maximum centripetal acceleration is lowered with respect to the case of vanishing forces at the tires.

In Figure 11.38, the effect on handling diagram of an unequal distribution of the longitudinal forces at the tires of an axle is highlighted. Such an unequal distribution generates the moments M_{zF} and M_{zR} shown in Figure 11.35. If the moments are positive as shown in Figure 11.38, the understeering

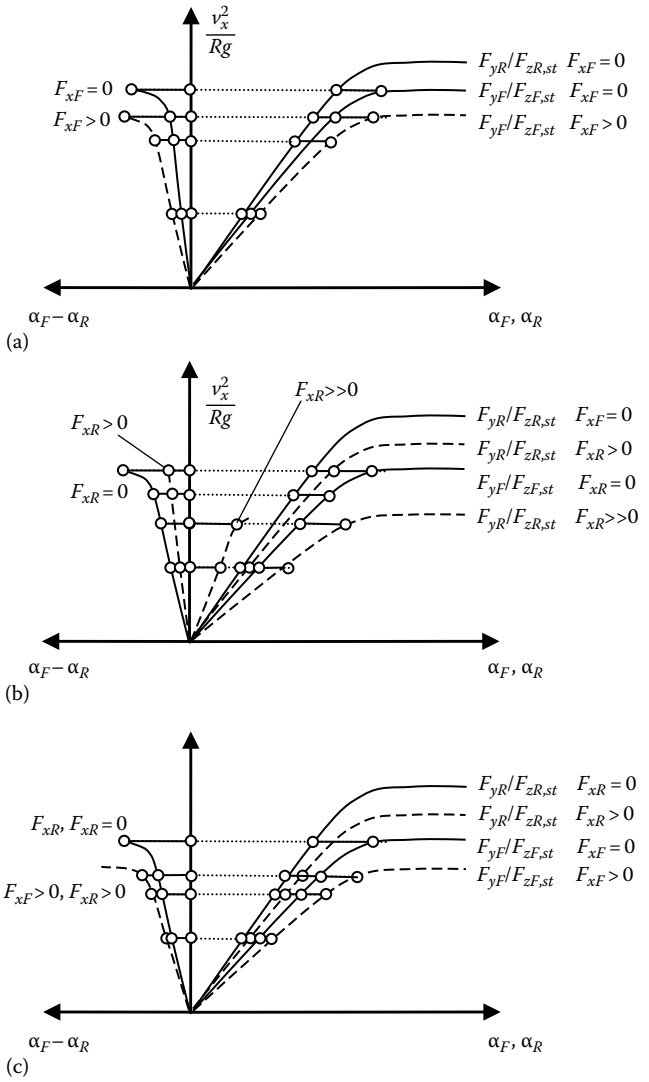


FIGURE 11.37 Effects on the handling diagram of different driveline architectures. (a) Front-wheel drive. (b) Rear-wheel drive. (c) All-wheel drive.

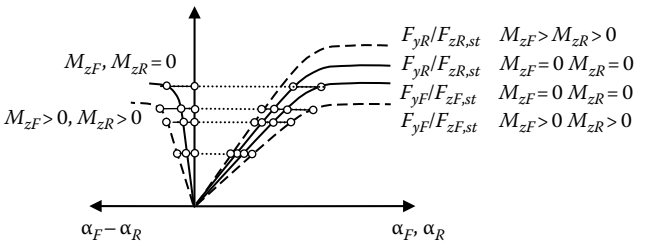


FIGURE 11.38 Effects on the handling diagram of torque vectoring, that is, unequal distribution of longitudinal forces at the two tires of the same axle (the moments M_{zF} and M_{zR} are generated, respectively, at the front and rear axles).

gradient is increased with respect to the case in which $M_{zF} = M_{zR} = 0$. Obviously, it is not necessary that both of the two moments M_{zF} and M_{zR} are to be applied. Again, the attainable maximum centripetal acceleration is lowered with respect to the case of vanishing vertical moments at the axles.

Rear-wheel-drive cars exhibit an oversteering character (see Figure 11.37b), so they get advantage from a torque vectoring like the one depicted in Figure 11.38 that counteracts oversteering with understeering [87–89].

The rolling resistances can be very high during cornering (see footnote; Section 11.3.4). Such a rise of rolling resistance can be introduced by considering the effect on tire characteristic as shown in this section. The aerodynamic drag can be introduced as mentioned for rolling resistance.

In this section, the effect of lateral load transfer has been neglected; in the following section, such an effect is dealt with.

11.4.1.4 Suspension Kinematics and Elastokinematics

In [8], Pacejka has given an inspiring synthesis on the influence of vehicle parameters (depicted in Figure 11.39) on effective axle characteristic (Figure 11.32). Such a contribution is summarized here. The theoretical analysis is linear and limited to low lateral slips. Such analysis is anyway informative to give an idea on the complexity of the handling behavior depending on axle kinematics and elastokinematics at higher lateral slips. Another similar contribution can be found in [2]. The axle characteristics (Figures 11.32 and 11.33) at low lateral slips read (symbols are reported in Figure 11.39)

$$\frac{F_{yi}}{F_{zi}} = \frac{F_{yi,in} + F_{yi,out}}{F_{zi,in} + F_{zi,out}} = \frac{C_i \alpha_{ai}}{F_{zi}} \quad \left(i = F, RF = \text{front}, R = \text{rear}, in = \text{inner wheel}, out = \text{outer wheel}\right)$$

$$C_i = \frac{2C_{\alpha io}}{1 + A + B - C + D}$$

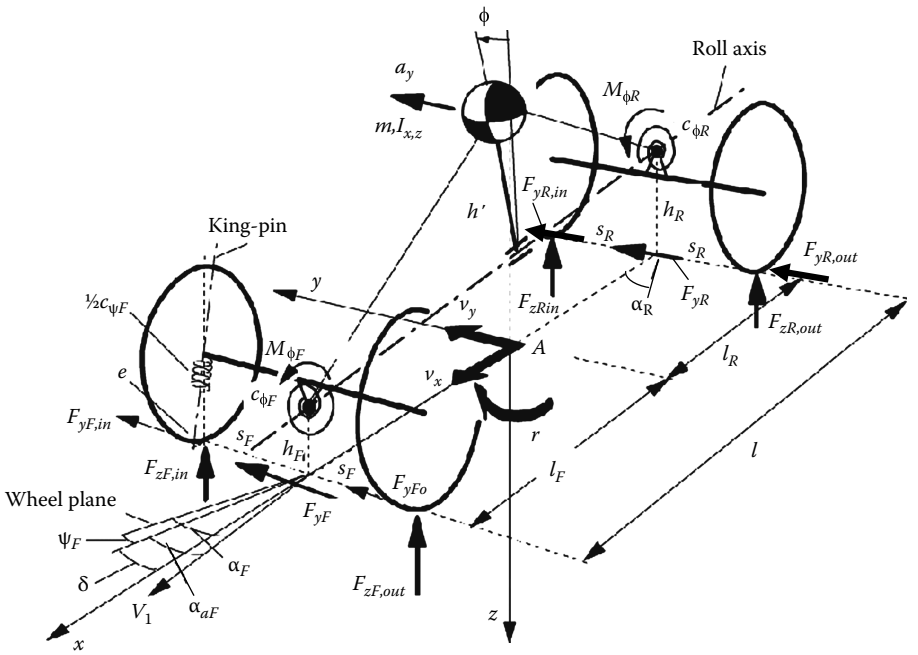


FIGURE 11.39 Vehicle model for cornering behavior. (Adapted from Pacejka, H., *Tyre and Vehicle Dynamics*, SAE International, Warrendale, PA, 2005.)

$$A = \frac{2l\left(\varepsilon_i C_{\alpha io} + \tau_i C_{\gamma io}\right)h'}{\left(l - l_i\right)\left(c_{\phi F} + c_{\phi R} - mgh'\right)}$$

$$B = \frac{2C_{\alpha io}\left(e_i + t_i\right)}{c_{\psi i}} \tag{11.64}$$

$$C = 2C_{\alpha io}c_{sfi}$$

$$D = \frac{2l\sigma_i}{l - l_i}\left(\zeta_{\alpha io}\psi_{io} + \zeta_{\gamma io}\gamma_{io}\right)$$

where

α_{ai} ($i = F, R$; Figure 11.39) represents the lateral slip that the axle would have if ψ_i were zero

$2C_{\alpha io}$ is obtained from the sum of the outer tire cornering stiffness (subscript *out*) plus the inner tire cornering stiffness (subscript *in*) $C_{\alpha i} = C_{\alpha i,\,out} + C_{\alpha i,\,in}$

The tire cornering stiffness depends on the vertical force (see Figure 11.33):

$$C_{\alpha i,out} = C_{\alpha io} + \zeta_{\alpha i}\left(F_{zi,out} - F_{zio}\right) = C_{\alpha io} + \zeta_{\alpha io}\Delta F_{zi}$$

$$C_{\alpha i,in} = C_{\alpha io} - \zeta_{\alpha i}\left(F_{zio} - F_{zi,in}\right) = C_{\alpha io} - \zeta_{\alpha io}\Delta F_{zi}$$

Finally,

$$C_{\alpha i,out} + C_{\alpha i,in} = 2C_{\alpha io}$$

A is related to roll steer (i.e., the steering induced by body roll) and to camber induced by body roll. The steering angle at the two wheels of an axle due to body roll can be linearized as

$$\psi_r = \varepsilon_i \phi$$

The camber angle at the two wheels of an axle due to body roll can be linearized as

$$\gamma_r = \tau_i \phi$$

(the body roll can be computed as $\phi = -ma_y h'/c_{\phi F} + c_{\phi R} - mgh'$ where $c_{\phi F}$ and $c_{\phi R}$ are, respectively, the front and rear roll stiffnesses; see Section 11.3.2.2)

$C_{\gamma io}$ is the axle camber stiffness depending on tire camber stiffness ($F_y = C_\gamma \gamma$), defined as $C_{\gamma i} = C_{\gamma io} + \zeta_{\gamma io}\Delta F_{zi}$.

Other symbols l, l_i, h', m are depicted in Figure 11.39, and their meaning is readily understandable.

B is related to steering compliance that causes the steering of the wheels due to the application of a lateral force, e_i represents the caster arm (Figure 11.39), t_i represents the trailing arm [8], and $c_{\psi i}$ represents the steering stiffness (rotation of the wheel around its steering axis; see Figure 11.39).

C is related to the suspension compliance that causes the steering of the wheels due to the application of a lateral force; the steering angle is

$$\psi_{sfi} = c_{sfi_i} F_{yi}$$

c_{sfi} represents the rotational stiffness of the suspension in lateral direction.

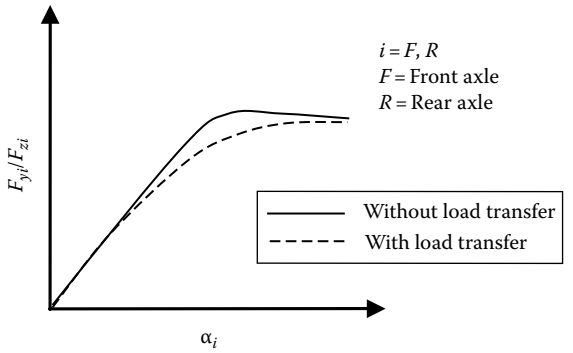


FIGURE 11.40 Effect of load transfer on effective axle characteristic.

D is related to the load transfer effect (at an axle, during steady-state cornering, the vertical force at the outer wheel is higher than the vertical force at the inner one). The vertical force at respective outer and inner wheel is

$$F_{zi,out} = \frac{F_{zi}}{2} + \Delta F_{zi}$$

$$F_{zi,in} = \frac{F_{zi}}{2} - \Delta F_{zi}$$

where F_{zi} is the static vertical load at the ith axle

$$\Delta F_{zi} = \sigma_i m a_y$$

$$\sigma_i = \frac{1}{2s_i}\left(\frac{c_{\phi i}}{c_{\phi F} + c_{\phi R} - mgh'}h' + \frac{l - l_i}{l}h_i\right)$$

ψ_{io} and γ_{io} refer, respectively, to the static toe angle and static camber angle.

Equation 11.64 states that the effective axle characteristic at low lateral slip angle is influenced by many parameters of the vehicle. The influence of the wheelbase and of the center of gravity position is mixed up with the influence of other parameters, so a simple synthesis is not possible. The designer can easily change the axle characteristic combining the many parameters appearing in Equation 11.64. The effect of load transfer is present only if ψ_{io} and γ_{io} are not null. The effect of load transfer is substantial at medium lateral slips (Figure 11.40). More details on load transfer and elastokinematics effects of effective axle characteristic can be found in [8].

11.4.2 *Rollover*

An excellent overview of the main issues referring to rollover has been given in [7,111]. Figure 11.41 shows a vehicle running at steady state into a bend, without suspension and for which both of the two inner wheels miss the contact with the ground at the same time; the rollover threshold is $a_y/g = s/h$, where s is half the track width, h is the height of the center of gravity from the ground, m is the vehicle mass, and a_y and g are the lateral acceleration and the gravitational acceleration, respectively. Rollover threshold is in the range 1.2–1.7 for cars, in the range 0.9–1.1 for pickup truck, and in the range 0.4–0.8 for trucks.

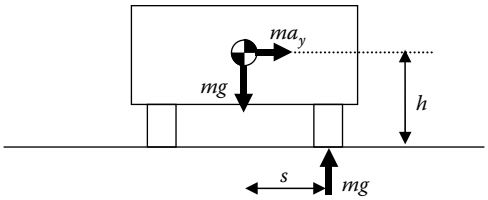


FIGURE 11.41 Simple vehicle model describing rollover initiation.

The actual vehicle is more similar to the one shown in Figure 11.39 than the one in Figure 11.41. In the actual case, during rollover initiation only one of the four wheels leaves the ground. In this case, the rollover threshold is obtained observing (Figure 11.39) that the vertical force at the inner wheel of the ith axle is $F_{zi,\,in} = F_{zi}/2 - \Delta F_{zi} = 0$ where $F_{zi} = mg(l - l_i)/l$ is the static vertical load at the ith axle and $\Delta F_{zi} = \sigma_i m a_y$ is the load transfer at the ith axle, with $\sigma_i = 1/2 s_i((c_{\phi i}/c_{\phi F} + c_{\phi R} - mgh')h' + (l - l_i/l)h_i)$. Thus, the rollover threshold is

$$\frac{a_y}{g} = \frac{s_i}{h_i + \dfrac{c_{\phi i}}{c_{\phi F} + c_{\phi R} - mgh'} \dfrac{l}{l - l_i} h'} \quad \left(i = F, R \; F = \text{front}, R = \text{rear}\right) \tag{11.65}$$

Notice that $h_i + h' \cong h$ (Figures 11.39 and 11.41). Such a threshold is lower (about 5%) than the one previously defined (s/h).

Rollover during transient maneuvers has been studied by simulating a step steer input by means of a model like the one shown in Figure 11.39. A step steer maneuver that produces at its end a lateral acceleration level (a_y/g) below the rollover threshold defined by Equation 11.65 can cause rollover, because of the body roll overshoot. Such an end lateral acceleration level depends on the body roll damping ratio and can be lower by a factor of about 10% to the previously derived rollover threshold.

Under a sinusoidal steer input, the body roll can reach its resonance; in this case, a wheel can leave the ground at a lateral acceleration (at the body center of gravity, a_y/g) lower by a factor 20% with respect to the rollover threshold defined by Equation 11.65. A sinusoidal steer input attaining the lateral acceleration needed for rollover can be hardly reproduced in reality, and so it is of little significance. The coupling of yaw motion with body roll motion somehow moderates the peak lateral acceleration.

The tipping phenomenon is another cause for rollover. Obviously the level of energy dissipated during the impact is important to reduce the energy required for rollover (rotational kinetic energy plus potential energy to raise the center of gravity). The s/h threshold appears to be important in the case of tipping too.

In [7], the analysis of rollover accidents is summarized both for cars and for trucks. It appears that not only the threshold s/h is important but also a number of other factors contribute equally to the rollover occurrence.

Other contributions on rollover are given in [90–99].

11.5 VEHICLE INERTIA PROPERTIES AND DIMENSIONS

The inertia properties of a rigid body are a set of 10 parameters, namely, the mass (1 parameter), the location of the center of mass (3 parameters), and the inertia tensor (6 parameters, usually the inertia tensor is related to the center of mass). The vehicle is composed by a number of bodies (subsystems) that can be considered rigid in the frequency range 0–25 Hz. So the inertia properties of a vehicle refer actually to the ones of each single vehicle subsystem.

The inertia properties of the body of a vehicle depend on design choices. Some design targets are obvious (minimum mass, minimum height of the center of mass). The targets referring to the

TABLE 11.6
Mass Distribution in Cars of Different Sizes

Components	Sub Compact [kg]		Compact [kg]		Upper Mid-Range [kg]	
Body in white	251,070	31.2%	313,083	28.0%	386,446	25.7%
Suspension	135,620	16.9%	193,314	17.3%	214,682	14.3%
Engine	71,683	8.9%	97,963	8.8%	178,242	11.8%
Interiors	36,695	4.6%	67,649	6.1%	85,881	5.7%
Seats	43,638	5.4%	61,555	5.5%	65,992	4.4%
Paintings and similar	33,578	4.2%	57,656	5.2%	73,759	4.9%
Cables and electric system *i*	39,722	4.9%	57,363	5.1%	105,142	7.0%
Brakes	27,484	3.4%	35,007	3.1%	50,900	3.4%
Exteriors	12,422	1.5%	32,297	2.9%	34,786	2.3%
Auxiliaries	24,076	3.0%	31,726	2.8%	72,504	4.8%
Windscreen and windows	29,734	3.7%	30,248	2.7%	37,710	2.5%
Intake/exhaust syst.	18,842	2.3%	24,989	2.2%	44,501	3.0%
Dashboard	13,418	1.7%	22,876	2.0%	21,188	1.4%
Mobile exteriors	21,866	2.7%	22,382	2.0%	33,213	2.2%
Controls	15,984	2.0%	18,825	1.7%	14,442	1.0%
Tank and related system	6,083	0.8%	12,143	1.1%	15,712	1.0%
Steering system	5,220	0.6%	9,628	0.9%	11,460	0.8%
Coatings	1,356	0.2%	8,974	0.8%	11,543	0.8%
Air conditioning/heating	9,006	1.1%	8,696	0.8%	31,452	2.1%
Clutch	3,186	0.4%	5,656	0.5%	5,592	0.4%
Cooling system	3,340	0.4%	4,147	0.4%	9,114	0.6%
Total	804		1,116		1,504	

Source: Torre, S., The car weight (in Italian), *Autotecnica*, XXI, 9–10, 2002.

inertia tensor are less obvious; some researches have been made on this topic [100–102], but still an in-depth analysis is missing.

The number of parts in a car is the order of 10^4; Table 11.6 gives the mass of main car subsystems. It is clear that to decrease the mass, almost every single part has to be made lighter; actually, more than half part of the mass is given by a plethora of relatively light parts. The only heavy subsystems are the body structure and the suspension system [103].

The relationship between the mass and the wheelbase of today's cars can be taken as linear. In Figure 11.42, the mass and the wheelbase are plotted for European cars sold on the market (year 2008); the linear interpolation gives $m = m(l) = k_{ml}l + q$, where m is the mass, l is the wheelbase, and k_{ml} and q are constants ($k_{ml} = 2042$ kg/m and $q = -3971$ kg).

Given a car or a van, as the mass m increases, the center of gravity moves toward the rear axle; more generally, the position of the center of gravity is a function of the mass m. Given the horizontal distance of the center of gravity from the fore axle as a fraction of the wheelbase, that is, $l_F = k_l l$, the relation between k_l and m is (see, e.g., [84])

$$\Delta k_l = k_m \Delta m \ (k_m \text{ can be set with good accuracy equal to } 2 \times 10^{-4}\ \text{kg}^{-1})$$

or

$$k_l - k_{l0} = k_m \left(m - m_0 \right) \tag{11.66}$$

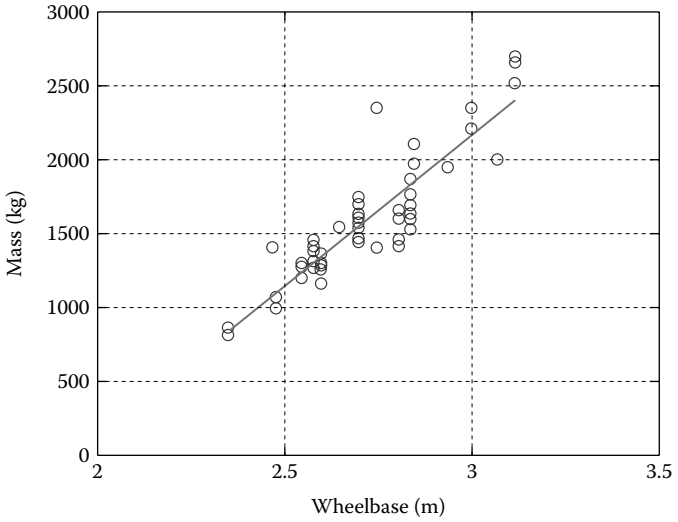


FIGURE 11.42 Relationship between mass (m [kg]) and wheelbase (l [m]) of European cars (as at year 2008). The linear interpolation gives $m = 2042l - 3971$ kg.

where k_{l0} and m_0 refer, respectively, to the center of gravity position and mass of a given unladen vehicle (for cars, k_{l0} is approximately in the range from 0.35 to 0.45, for truck and SUVs, in the range from 0.40 to 0.47).

Given a reference frame as in Figure 11.43, the center of mass R and the inertia tensor J (referred to the origin of the coordinate frame) are defined, respectively, as

$$R = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} \dfrac{\int_V \rho(r)\, x dV}{\int_V \rho(r) dV} \\ \dfrac{\int_V \rho(r)\, y dV}{\int_V \rho(r) dV} \\ \dfrac{\int_V \rho(r)\, z dV}{\int_V \rho(r) dV} \end{bmatrix}; \quad r = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

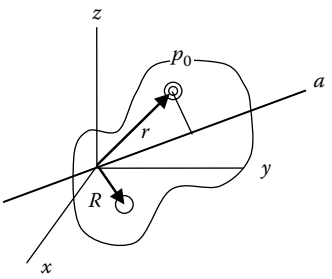


FIGURE 11.43 Reference coordinate frame for defining the inertia properties of a body; p_0 is a generic point of the body, identified by vector r. R identifies the location of the center of mass. The axis a identifies a generic axis of rotation.

$$\boldsymbol{J} = \begin{bmatrix} J_{xx} & J_{xy} & J_{xz} \\ & J_{yy} & J_{yz} \\ simm & & J_{zz} \end{bmatrix} = \int_V \left(\begin{bmatrix} y^2+z^2 & -xy & -xz \\ -yx & x^2+z^2 & -yz \\ -zx & -zy & x^2+y^2 \end{bmatrix} \right) \rho(r) dV$$

where $\rho(r)$ is the density of the body at the point defined by vector r.

The relationship between the moment of inertia around the axis $a(J_a)$ (Figure 11.43) and the inertia tensor $\boldsymbol{J}$ is $J_a = \nu^T J \nu$,where ν is the vector of the direction cosines of the line a.

Due to the aforementioned definition of $\boldsymbol{J}$, the following relationships hold among the components of the inertia tensor $\boldsymbol{J}$ [104]:

$$J_{xx} \geq 0, \quad J_{yy} \geq 0, \quad J_{zz} \geq 0$$

$$J_{xx} + J_{yy} \geq J_{zz}, \quad J_{xx} + J_{zz} \geq J_{yy}, \quad J_{yy} + J_{zz} \geq J_{xx}$$

$$A^2 \leq 1, \quad B^2 \leq 1, \quad C^2 \leq 1$$

$$A^2 + B^2 + C^2 - 2 \cdot A \cdot B \cdot C \leq 1$$

where

$$A = \frac{2|J_{xy}|}{\sqrt{J_{zz}^2 - \left(J_{xx}^2 - J_{yy}^2\right)}}, \quad B = \frac{2|J_{xz}|}{\sqrt{J_{yy}^2 - \left(J_{xx}^2 - J_{zz}^2\right)}}, \quad C = \frac{2|J_{yz}|}{\sqrt{J_{xx}^2 - \left(J_{zz}^2 - J_{yy}^2\right)}}$$

Thus, the six components of the inertia tensor *cannot* take whichever value.

In [84], an extensive survey of measured inertia properties of wheeled ground vehicles is given. The center of gravity height (C_G) divided by the roof height (R_H) is found within the range $0.36 \leq C_G/R_H \leq 0.42$. C_G/R_H increases very slightly as function of the vehicle mass, either for cars, vans, trucks, or SUVs. For cars, the ratio a/W_B (a horizontal distance between the center of gravity and the front axle axis; W_B wheelbase) can be found in the range 0.36–0.40 and raises to the range 0.45–0.49 when the vehicle is fully loaded. In other words, the center of gravity is moved sensibly toward the rear axle when the vehicle is fully loaded.

Relevant statistics among vehicle dimensions and main diagonal inertia tensor components are given in [100] (Table 11.7); in particular, the values of the main diagonal inertia tensor components J_{xx}, J_{yy}, J_{zz} are given as function of the wheelbase W_B, the track width T_W, the roof height $R_H = H_r$, and the weight W_t (the prediction error can be rather large, ±25%, and in some specific cases even more):

$$\begin{bmatrix} J_{xx} \\ J_{yy} \\ J_{zz} \end{bmatrix} = \begin{bmatrix} 10^{k_{1x}} \cdot W_b^{k_{2x}} \cdot T_w^{k_{3x}} \cdot H_r^{k_{4x}} \cdot W_t^{k_{5x}} \\ 10^{k_{1y}} \cdot W_b^{k_{2y}} \cdot T_w^{k_{3y}} \cdot H_r^{k_{4y}} \cdot W_t^{k_{5y}} \\ 10^{k_{1z}} \cdot W_b^{k_{2z}} \cdot T_w^{k_{3z}} \cdot H_r^{k_{4z}} \cdot W_t^{k_{5z}} \end{bmatrix} \tag{11.67}$$

Coefficients of Equation 11.67 (Table 11.6)

$k_{1x} = -2.14 \cdot k_{2x} = -0.16 \cdot k_{3x} = 1.94 \cdot k_{4x} = 0.36 \cdot k_{5x} = 0.94$
$k_{1y} = -2.00 \cdot k_{2y} = 1.53 \cdot k_{3y} = 0.25 \cdot k_{4y} = 0.10 \cdot k_{5y} = 1.02$
$k_{1z} = -1.78 \cdot k_{2z} = 1.43 \cdot k_{3z} = 0.38 \cdot k_{4z} = 0.02 \cdot k_{5z} = 0.98$

TABLE 11.7
Correlation Coefficients among Vehicle Dimensions and Main Diagonal Components of the Inertia Tensor

	Log W_B	Log T_W	Log R_H	Log W_t	Log I_{xx}	Log I_{yy}	Log I_{zz}
Log W_B	1.0000						
Log T_W	0.7568	1.0000					
Log R_H	0.3811	0.5125	1.0000				
Log W_t	0.6639	0.7121	0.6462	1.0000			
Log I_{xx}	0.6793	0.8208	0.6805	0.9231	1.0000		
Log I_{yy}	0.8521	0.7952	0.5953	0.9242	0.9086	1.0000	1.0000
Log I_{zz}	0.8571	0.8052	0.5860	0.9268	0.9056	0.9942	

Source: From Allen, W.R., Estimation of passenger vehicle inertial properties and their effect on stability and handling, SAE Paper No. 2003-01-0966.
Wheelbase, W_B; track width, T_W; roof height, R_H; and weight, W_t.

Very simple and approximate formulae to estimate J_{xx}, J_{yy}, and J_{zz} are, respectively,

$$J_{xx} = \frac{M \cdot s^2}{2}$$

$$J_{yy} \approx J_{zz}$$

$$J_{zz} = M \cdot a \cdot b \text{ or } J_{zz} = k \cdot M' W_b^{3/2}$$

where
M is the mass of the body
s is the track width
a and b are, respectively, the horizontal distances of the center of gravity from the front and rear axles
k is constant (that can be obtained from Equation 11.67)
M' is the mass of the vehicle
W_b is the wheelbase

Under the hypothesis that the powertrain is rigidly connected to the car body, cars with a front engine feature usually a non-negligible J_{xz} [84,101,102]. This occurrence can couple the yaw and roll motions; to alleviate this, the roll axis can be sloped toward the front [2].
In [101], the influence of the inertia tensor on ride and handling of a car was highlighted. Small variations of the components of the inertia tensor may produce significant variations of ride comfort and on center handling of vehicles [2,8]. The inertia tensor plays also a very important role during panic evasive maneuvers involving the dynamic stability of road vehicles.

It is rather hard to compute the inertia properties of ground vehicles (actually there is a huge number of parts to take into account; see Table 11.6), so it is convenient to *measure* them. Also the measurement is not simple and requires special skills [104]. In [84,105,106], an overview of existing methods and facilities to measure the inertia properties of vehicles is given.

In [101,107], the needed measurement accuracy of inertia properties was proposed in order to perform reliable simulations, either for full vehicles or a subsystem such as the engine. Referring to the car body, the mass should be known with an uncertainty <0.5%; the main diagonal components of



FIGURE 11.44 An accurate system for measuring the inertia properties of cars. (From Mastinu, G. et al., A new test rig for the identification of the inertia tensor of a rigid body, SAWE Paper 3315.)

the inertia tensor should be measured with a tolerance <3% and the extra diagonal components within the uncertainty range ±15 kg m^2 (Figure 11.44). Referring to the system engine + gearbox of a car, the appropriate computation of both the natural frequencies and the forces transmitted to the chassis should be performed when the main diagonal components of the inertia tensor are known with an uncertainty <2.5%; the position of the center of gravity is measured with an uncertainty <3 mm; the mass should be known with an uncertainty <1.5%.

11.6 CONCLUSION

The wheel has been invented about 5000 years ago. Employing wood has allowed to shape the wheel from a log-like cylinder to a disk with a relatively low mass and high diameter to run on obstacles. Since the beginning, the vehicles were fitted with wheel couples (i.e., axles, two wheels rotating around the same axis), and two-axle vehicles were used. The track of ancient Roman carts was standardized (possibly the cart track width was related to the cross width of horses), and today's road vehicles have approximately the same track of such ancient vehicles. Extensive implementation of brakes, steering, and suspensions was firstly introduced by Romans.

The dimensions of today's cars are strictly related to anthropometry, customer requirements, and technical equipment. The posture of humans inside the car largely influences the vehicle geometry. However, there is not a well-accepted rule to set the wheelbase.

In non-tilting vehicles, the best way to avoid rollover during cornering and braking is to arrange four wheels at the four corners of the car body.

In order to describe the running of vehicles on irregular road profile, the so-called quarter vehicle model (2 degrees of freedom [d.o.f.]) is very useful; such a simple model can be employed for conceptual design purposes under broad conditions. Ride (dis)comfort, road holding, and working space (rattling of the wheel with respect to body) are the output performance indices of the quarter vehicle model. Pitch body motions can be somehow taken into account by using a couple of the quarter-car models.

A vehicle with a single suspension arrangement (i.e., a simply sprung vehicle whose body is resting on stiff wheels and springs like ancient Roman carts) is not efficient to filter properly the excitation produced by the irregular road profile. Actually, for a simple quarter-car model with 1 d.o.f. excited by a standard irregular road profile, both the variance of the body mass vertical acceleration (discomfort index) and the variance of the vertical force at the wheel–ground interface (road holding index) are unbounded. Providing an additional suspension system (pneumatic tire) and referring to a 2 d.o.f. quarter-car model, the mentioned variances are made bounded, that is, adding a

pneumatic tire to a simply sprung vehicle makes a big difference, both from a theoretical and from a practical point of view. Actually the tire acts as low-pass filter. Two is the minimum number of passive suspension stages to obtain a bounded response of a ground vehicle to common random road irregularity. Such conclusions on suspension system behavior have been obtained by deriving relatively simple analytical formulae.

Providing a vertical massless spring between the vehicle and the ground is necessary (but not sufficient) to obtain a bounded response of the passively suspended vehicle to road irregularity excitations. The pneumatic tire represents the solution to such a technical need.

The working principle of the pneumatic tire surpasses the one of the solid wheel in that the Hertz contact pressure depending on wheel radius is no more the leading phenomenon (if the pneumatic tire were a thin-walled membrane, the contact pressure were simply the inner pressure). The load-carrying capacity of the tire is related not only to inner pressure but also to geometry. With reasonable approximation for concept design, the pneumatic tire load-carrying ability increases both with the square root of the wheel radius and linearly with inner pressure and cross width. The same wheel parameters affect in the same way the pneumatic tire stiffness. Such a synthesis on tire behavior has been obtained by deriving relatively simple analytical formulae.

Inherently, non-perfectly stiff wheels are prone to rolling resistance. Actually, hysteresis implies a moment that counteracts the rotation of the wheel. Such conclusion has been derived by analyzing a simple model featuring a wheel with an elastic tire. There are many different phenomena that lead to rolling resistance, involving both vertical and longitudinal tire compliance.

Rolling resistance is important but not the most important drag force during urban running conditions when inertia resistance is prevailing. Due to this, the powertrain of vehicles delivers often the highest power during acceleration. An analysis based on the NEDC has been introduced. Minimizing the energy to drive the NEDC requires reducing the vehicle mass and, to a less extent, the tire rolling resistance and aerodynamic drag. The recuperation of the kinetic energy does not lead to savings higher than about 20%.

The architecture of the driveline of two-axle road vehicles (front-, rear-, or all-wheel drive) depends—among other relevant aspects—on longitudinal acceleration performance (power on). The optimal torque distribution between the front and rear axles should be tuned (as for braking) as function of the longitudinal acceleration.

The active safety (handling) performance of a two-axle vehicle depends mostly on tire characteristics, inertial and geometric data (mass, center of gravity location, wheelbase), drive axle/torque vectoring, and suspension kinematics (or elastokinematics). All of these factors affecting vehicle architecture can be taken into account by the handling diagram theory. Actually the handling diagram theory allows an insight into the nonlinear phenomena involved within handling and allows to quickly understand the relation between vehicle architecture and vehicle performance. The crucial role of pneumatic tires on handling has been highlighted; actually the relationship between steering angle and centripetal acceleration is directly related to tire characteristics.

In order to investigate the influence on handling of vehicle inertial data and geometry, to simplify the theoretical investigation, the cornering stiffness of tires has been considered constant or linearly dependent on tire vertical load. In this last case, the understeering gradient is insensitive both to the vehicle mass and to the location of the center of gravity. Additionally, in this case, the yaw velocity gain and the damping ratio (of both the yaw rate and the sideslip angle responses to a steering input) do not depend on the location of the center of gravity. On the contrary, if the tire cornering stiffness characteristic is independent on tire vertical load (i.e., constant), the vehicle dynamic performance is very sensitive to all of its geometric and inertial parameters. Such conclusions have been derived by simple analytical formulae.

The effects on handling of drivetrain architecture (front-, rear-, or all-wheel drive) become evident taking into account how the handling diagram is constructed. Always traction forces reduce the maximum centripetal acceleration attainable, and the understeering is, respectively, increased or decreased by front- or rear-wheel drive. Torque distribution affects the handling performance in a way that is easily predicted by a simple and original formula derived in the chapter.

The kinematics and elastokinematic effects on the handling diagram (and on active safety) have been considered with reference to both the low and high levels of centripetal acceleration. At low levels of centripetal acceleration, the influence of lateral load transfer is generally vanishing; at high levels of centripetal acceleration, every single parameter of the suspension system may play an important role in defining the stability of a vehicle.

The inertia properties (mass, center of mass location, and body inertia tensor) are of crucial importance to assess accurately the dynamic behavior of vehicles. Some inertia properties are related to either vehicle geometry or among themselves. The mass is statistically linearly related to the wheelbase; the center of gravity horizontal distance from the front axle increases linearly with the vehicle mass as the vehicle is loaded. There are more or less simple relationships giving the main diagonal inertia tensor components as function of vehicle geometry. The experimental measurement of vehicle inertia properties is relatively quick and accurate.

After the analyses performed in this chapter, a comprehensive set of basic relations has been given to highlight how the vehicle architecture is related to vehicle dynamic performance.

The main issues referring to the conceptual design of road vehicles related to dynamics have been addressed referring to sustainable technology available today. In the future, the paradigms on which vehicle is conceived could change even abruptly provided that new technologies are made available and applied. An investigation on far future road vehicle conceptual design was out of the scope of this chapter.

REFERENCES

1. Genta, G., *Motor Vehicle Dynamics: Modelling and Simulation*, Levrotto & Bella, Turin, Italy, 1997.
2. Mitschke, M., Wallwntowitz, H., *Dynamik der Kraftfahrzeuge*, Springer, Berlin, Germany, 2004.
3. Braess, H.H., Seiffert, U.W., *Handbook of Automotive Engineering*, SAE, Warrendale, PA, 2005.
4. Bosch, *Automotive Handbook*, SAE, Warrendale, PA, 2005.
5. Morello, L., Pia, G., Rosti Rossini, L., Tonoli, A., *La Carrozzeria Progetto dei componenti*, ATA, Turin, Italy, 2009.
6. Morello, L., Pia, G., Rosti Rossini, L., Tonoli, A., *La Carrozzeria Progetto del sistema*, ATA, Turin, Italy, 2009.
7. Gillespie, T.D., *Fundamentals of Vehicle Dynamics*, Society of Automotive Engineers, Warrendale, PA, 1992.
8. Pacejka, H., *Tyre and Vehicle Dynamics*, SAE International, Warrendale, PA, 2005.
9. French, T., *Tyre Technology*, Adam Hilger, Bristol, U.K., 1989.
10. Clarks, S. (Ed.) *Mechanics of Pneumatic Tyres*, U.S. Department of Transportation (DOT), National Highway Traffic Safety Administration (NHTSA), Washington, DC, 1980.
11. Popp, K., Schiehlen, W.O., *Ground Vehicle Dynamics*, Springer, Berlin, Germany, 2010, ISBN 978-3-540-24038-9.
12. Genta, G., Morello, L., *L'autotelaio. Progetto dei sistemi*, ATA (Association of Italian Automotive Engineers), Turin, ISBN 978-88-8218-129-1, 2007.
13. Genta, G., Morello, L., *L'autotelaio. Progetto dei componenti*, ATA (Association of Italian Automotive Engineers), Turin, ISBN 978-88-8218-129-1, 2007.
14. Dixon, J.C., *Tires, Suspension and Handling*, Society of Automotive Engineers, Warrendale, PA, 1996.
15. Howard, G., Whitehead, J.P., Bastow, D., *Car Suspension and Handling* (Fourth Edition), SAE, Warrendale, PA, 2004.
16. Stone, R., Ball, K., *Automotive Engineering Fundamentals*, SAE, Warrendale, PA, 2004, ISBN 978-0-7680-0987-3.
17. Garrett, T.K., Steeds, W., Newton, N., *The Motor Vehicle*, Elsevier Butterworth-Heinemann, Oxford, U.K., 2001.
18. Gent, A., Walter, D., The pneumatic tire, U.S. Department of Transportation (DOT), NHTSA, Washington, DC, 2006.
19. Singer, C., *A History of Technology*, University of Michigan Library, Scholarly Publishing Office, Ann Arbor, MI, 2008.
20. Weller, J., Roman traction system, http//www.humanist.de/rome/rts/index.html

21. Daumas, M. (Ed.), *Histoire Général des Techniques*, Presse Universitaires de France, Paris, France, 1962.
22. Ellis, C.H., *The Lore of the Train*, AB Nordbok, Gothenburg, Sweden, 1975.
23. Derry, K., Williams, T.I., *A Short History of Technology*, Clarendon Press, Oxford, U.K., 1960.
24. www.lexiqueducheval.net
25. White, K.D., *Greek and Roman Technology*, Cornell University Press, Ithaca, NY, 1984.
26. Landels, J.G., *Engineering in the Ancient World*, California University Press, Berkeley, CA, 2000.
27. Röring, C.W., *Untersuchungen zu römischen Reisewagen*, Numismatischer Verlag G.M. Forneck, Koblenz, 1983.
28. Langdon, J., *Horse, Oxen, and Technological Innovation*, Cambridge University Press, Cambridge, U.K., 1986.
29. Weeber, K., *Alltag im alten Rom. Ein Lexikon*, Artemis & Winkler, Zürich, Switzerland, 1995.
30. McBurney, R., Drow, J., The elastic properties of wood, Forest Product Laboratory—Forest Service, U.S. Department of Agriculture, 1962.
31. http//www.pininfarina.com/index/storiaModelli/timeline/1960
32. Sharp, R.S., The stability and control of pivot framed tricycles, *Dynamics of Vehicles on Roads and Tracks*, Budapest, Hungary, Swets & Zeitlinger, Lisse, the Netherlands, 1984, pp. 564–577.
33. ISO 3833, Road vehicles—Types—Terms and definitions, 1977.
34. Seiffert, U.W., Braess, H.H., *Vieweg Handbuch Kraftfahrzeugtechnik*, Freidrich Vieweg & Sohn Verlagsgesellschaft mbH, Wiesbaden, Germany, 2001.
35. Bandow, F., Stahlecker, H., Deduction of main vehicle dimensions, *ATZ Automobiltechnische Zeitschrift*, 103(10), 8–11, 2001.
36. SAE Human Accommodation and Design Devices Stds. Comm., H-Point Machine (HPM-II)—Procedure for H-Point Determination—Benchmarking Vehicle Seats, Paper no. J4003, SAE, Warrendale, PA, 2005.
37. Mastinu, G. et al., Theoretical and experimental ride comfort assessment of a subject seated into a car, SAE Paper No. 2010010777, 2010.
38. SAE Human Accom. and Design Devices Stds. Comm., Positioning the H-point design tool—Seating reference point and seat track length, SAE Paper No. J4004, Warrendale, PA, 2005.
39. Verver, M.M., De Lange, R., Van Hoof, J., Wismans, J.S.H.M., Aspects of seat modelling for seating comfort analysis, *Applied Ergonomics*, 36(1), 33–42, 2005, Elsevier Ltd., Oxford, U.K.
40. Valero, B., Amirouche, F.M.L., Mayton, A., Jobes, C., Comparison of passive seat suspension with different configuration of seat pads and active seat suspension, SAE Paper No. 2007-01-0350, Warrendale, PA, 2007.
41. Philippart, N.L., Kuechenmeister, T.J., Stanick, J.M., Truck driver selected seat position model, SAE, Universal City, CA, 1986.
42. Park, S.J., Kim, C.B., The evaluation of seating comfort by the objective measures, SAE Paper no. 970595, 1997.
43. Rebiffé, R., The driving seat: Its adaptation to functional and anthropometric requirements, *Proceedings of the Symposium on Sitting Posture*, E. Grandjean (Ed.), London, U.K., pp. 132–147, 1969.
44. Grandjean, E. et al., VDT workstation design: Preferred setting and their effects, *Human Factors*, 25, 161–175, 1983.
45. Porter, J.M. et al., Beyond Jack and Jill: Design for individuals using HADRIAN, *International Journal of Industrial Ergonomics*, 33, 249–264, 2004.
46. Nagel, K., Comfort evaluation as ergonomic tool with application to interior concepts of vehicles, SAE Paper No.1999-01-1921, Warrendale, PA, 1999.
47. Dufour, F., Lino, F., Le Coz, J.Y., Vehicle accessibility, SAE Paper no. 2001-01-3431, Warrendale, PA, 2001.
48. Manary, M.A., Flannagan, C.A.C., Schneider, L., Development of an improved driver eye position model, SAE Paper No.980012, Warrendale, PA, 1998.
49. Ciuffi, L., Pignato, P., Pascuzzi, V., New ergonomic studies for transportation design, SAE Paper No. 2007-01-0344, Warrendale, PA, 2007.
50. Patwardhan, S.S., Bloebaum, C.L., Krovi, V.N., Using anthropometric modeling for optimal ergonomic considerations in automobile interior design, SAE Paper No. 2005-01-2718, Warrendale, PA, 2005.
51. Mergl, C., Bubb, H., Vogt, C., Kress, H., Ergonomic layout process for a driver working place in cars, SAE Paper No. 2006-01-2313, Warrendale, PA, 2006.
52. Bouchard, C., Aoussat, A., Modellization of the car design process, *International Journal of Vehicle Design*, 31(1), 1–10, 2002.

53. Lin, M.-C., Yan, I.-H., Development of a computer-assisted procedure for car style design, *International Journal of Vehicle Design*, 35(4), 289–306, 2004.
54. Hsiao, S.W., Fuzzy set theory applied to car style design, *International Journal of Vehicle Design*, 15 (3–5), 225–278, 1994.
55. Xiangdong, H., Chaolin, X., Yuecong, X., Intelligent car styling technique and system based on a new aerodynamic-theoretical model, *Chinese Journal of Mechanical Engineering*, 15(2), 110–115, 2002.
56. Hucho, W.H., *Aerodynamics of Road Vehicles*, SAE, Warrendale, PA, 1998.
57. Gleason, M., Han, T., Khalighi, B., Sumantran, V., Williams, J., On the use of reference models in automotive aerodynamics, SAE Paper No. 2004-01-1308, 2004.
58. Ding, W., Williams, J., Karanth, D., Sovani, S.D., CFD application in automotive front-end design, SAE Paper No. 2006-01-0337, Warrendale, PA, 2006.
59. Knoke, O., Pfaff, I., Winkler, J., Front final interpretation for the passive pedestrian protection, *ATZ Automobiltechnische Zeitschrift*, 108(11), 952–955, 2006.
60. Neal, M.O., Front structure design procedure for optimal pedestrian leg impact performance, *Proceedings of the ASME Design Engineering Technical Conference*, Vol. 1, 2004, American Society of Mechanical Engineers, New York, pp. 199–207.
61. Dodds, C.J., Robson, J.D., The description of road surface roughness, *Journal of Sound and Vibration*, 31, 175–183, 1973.
62. Parkihlowski, I.G., Investigations of the probability characteristics of the surfaces of distributed types of roads, *Avtomobi'naya Promyshlennost*, 8, 18–22, 1968.
63. Ammon, D., Problems in road surface modelling, *12th IAVSD Symposium*, Lyon, France, 1991.
64. Kamash, K., Robson, D.J., The application of isotropy in road surface modelling, *Journal of Sound and Vibration*, 57, 89–100, 1978.
65. Nigam, J., *Introduction to Random Vibrations*, The MIT Press, Cambridge, MA, 1983, ISBN-10 0-262-14035-7 ISBN-13 978-0-262-14035-5.
66. Venhovens, P., Optimal control of vehicle suspensions, PhD thesis, Delft University of Technology, Delft, the Netherlands, 1994.
67. Mueller, P.C., Schiehlen, W.O., *Linear Vibrations*, Martinus Nijohoff Publishers, Dordrecht, the Netherlands, 1985.
68. Gobbi, M., Mastinu, G., Analytical description and optimization of the dynamic behaviour of passively suspended road vehicles, *Journal of Sound and Vibration*, 245(3), 457–481, 2001.
69. Sharp, R., Crolla, D., Road vehicle suspension system design—A review, *Vehicle System Dynamics*, 16, 167–192, 1987.
70. Reimplell, J. et al., *The Automotive Chassis Engineering Principles*, Butterworth-Heinemann, Oxford, U.K., 2001.
71. Bourcier de Carbon, C., Theorie Mathématique et réalisation pratique de la suspension amortie des véhicules terrestres, *3rd Congrès Technique de l'Automobile*, Paris, France, October 1950.
72. Sharp, R.S. et al., Influences of suspension kinematics on pitching dynamics of cars in longitudinal manoeuvering, *Vehicle System Dynamics*, 33(suppl.), 23–26, 2000.
73. Mastinu, G., Fainello, M., Study of the pneumatic tyre behaviour on dry and rigid road by finite element method, *Vehicle System Dynamics*, 21, 143–165, 1992, http://www.informaworld.com/smpp/title~db=all~content=t713659010~tab=issueslist~branches=27-v271
74. Chengjian, F., Dihua, G., The quantitative analysis and experimental verification of the tire static enveloping model using experimental modal parameters, *Vehicle System Dynamics*, 44(9), 675–688, September 2006.
75. Mastinu, G., Symbolical computation of the radial stiffness of car tyres, Internal Report, Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy, December 2009.
76. Mastinu, G., Previati, G., Gobbi, M., Analytical computation of the radial stiffness of pneumatic tyres, *XXI IAVSD Symposium*, Manchester, U.K., August 2011.
77. Rhyne, T.B., Development of a vertical stiffness relationship for belted radial tires, *Tire Science and Technology*, 33(3), 136–155, 2005.
78. Kim, S., Kondo, K., Akasaka, T., Contact pressure distribution of radial tire in motion with camber angle, *Tire Science and Technology*, 28, 2–32, 2000.
79. Kagami, S., Akasaka, T., Shiobara, H., Hasegawa, A., Analysis of the contact deformation of a radial tire with camber angle, *Tire Science and Technology*, 23(1), 26–51, 1995, Tire Soc Inc., Akron, OH.
80. Guzzella, L., Sciarretta, A., *Vehicle Propulsion Systems*, Springer Verlag, Berlin, Germany, 2005, ISBN-10 3-540-25195-2.
81. Westbrook, M., *The Electric and Hybrid Electric Car*, SAE, Warrendale, PA, ISBN 978-0-7680-0897-5, 2001.

82. Jurgen, R.K., *Electric and Hybrid-Electric Vehicles—Fuel Cell Hybrid EVs*, SAE, Warrendale, PA, 2010, ISBN 978-0-7680-5721-8.
83. SAE, *Transmission and Driveline*, SP 2911, SAE, Warrendale, PA, 2010, ISBN 978-0-7680-3425-7.
84. Heydinger, G. et al., Measured vehicle inertial parameters-NHTSA's data through November 1998, SAE Paper No. 1999-01-1336.
85. Morello, L., *Progetto della trasmissione meccanica*, ATA, Turin, Italy, 2005.
86. Kortuem, W., Lugner, P., *Systemsdynamik und Regelung von Fahrzeugen*, Springer Verlag, Berlin, Germany, 1994.
87. Mastinu, G., Battistini, E., The influence of limited-slip differentials on the stability of rear-wheel-drive automobiles running on even road with dry surface, *International Journal of Vehicle Design*, 14(2/3), 166–183, 1993.
88. Frendo, F., Greco, G., Guiggiani, M., Critical review of handling diagram and understeer gradient for vehicles with locked differential, *Vehicle System Dynamics*, 44(6), 431–447, June 2006.
89. Tomari, T., Shibahata, Y. et al., Limited slip differential, Patent U.S. 6857982, September 2002.
90. Bernard, J. et al., Vehicle rollover on smooth surfaces, SAE Paper No. 891991, 1989.
91. Jones, I.S., Penny, M.B., Engineering parameters related to rollover frequency, SAE Paper No. 900104, Warrendale, PA, 1990.
92. Lew, J., Piyabongkarn, D., Grogg, J., Minimizing dynamic rollover propensity with electronic limited slip differentials, SAE Paper No. 2006-01-1279, Warrendale, PA, 2006.
93. Dahlberg, E., Stensson, A., The dynamic rollover threshold—A heavy truck sensitivity study, *International Journal of Vehicle Design*, 40(1–3), 228–250, 2006, Inderscience Enterprises Ltd., Geneva, Switzerland.
94. Brandse, J. et al., Methodology for simulation of rollover cases, SAE Paper No. 2006-01-0558, 2006.
95. Gillespie, T.D. et al., Roll dynamics of commercial vehicles, *Vehicle System Dynamics*, 9, 1–17, 1980.
96. Winkler, C.B. et al., *Rollover of Heavy Commercial Vehicles*, SAE SP R-004, SAE, Warrendale, PA, 2000, ISBN 978-0-7680-0626-1.
97. NHSTA, http//www.nhtsa.gov/cars/rules/rulings/rollfinal/index.html
98. Marimuthu, R.P., A study on SUV parameters sensitivity on rollover propensity, SAE Paper No. 2006-01-0795, 2006.
99. Fernandes, C., Duarte, M., Archetypal vehicle dynamics model for resistance rollover prediction, SAE Paper No. 2010-01-0715, 2010.
100. Allen, W.R., Estimation of passenger vehicle inertial properties and their effect on stability and handling, SAE Paper No. 2003-01-0966.
101. Mastinu, G. et al., The influence of the body inertia tensor on the active safety and ride comfort of road vehicles, SAE Paper No. 02IBECD-14, 2002.
102. Doi, T., Nozawa, T., A theoretical analysis concerning the influence of the angle of axes of inertia on vehicle dynamics, SAE Paper No. 2006-01-0798, 2006.
103. Komatsu, Y., Arai, T., Abe, H., Sato, M., Nakazawa, Y., Application of aluminum for automobile chassis parts, SAE Technical Paper Series, SAE, Warrendale, PA, 1991.
104. Society of Allied Weight Engineers, *SAWE Handbook*, SAWE, Los Angeles, CA.
105. Mastinu, G., Gobbi, M., Previati, G., A new test rig for the identification of the inertia tensor of a rigid body, SAWE Paper 3315, 2003.
106. Gobbi, M., Mastinu, G., Previati, G., A method for measuring the inertia properties of rigid bodies, *Mechanical Systems and Signal Processing*, 25(1), 305–318, 2011.
107. Eberard, P. et al., Providing the inertia properties of vehicles and their subsystems for virtual reality and mechatronics applications, SAWE Paper 3435, 2009.
108. Bekker, M.G., *Off-The Road Locomotion*, University of Michigan Library, Scholarly Publishing Office, Ann Arbor, MI, 1960.
109. Mastinu, G., Gaiazzi, S., Montanaro, F., Pirola, D., A semi-analytical tyre model for steady- and transient-state simulations, *Vehicle System Dynamics*, 27, 2–21, 1997, http://www.informaworld.com/smpp/title~db=all~content=t713659010~tab=issueslist~branches=27-v2727
110. Pau, M., Leban, B., Baldi, A., Ultrasonic measurements of contact area and pressure distribution of a pneumatic tire on a rigid surface, *Tire Science and Technology*, *TSTCA*, 36(1), 43–62, January–March 2008.
111. NHTSA, Rollover Prevention, 49 CFR Part 575, Docket No. NHTSA-2000-6859, RIN 2127-AC64, (http//www.nhtsa.gov/cars/rules/rulings/rollover/)
112. Eckermann, E., *World History of the Automobile*, SAE, Warrendale, PA, 2001.

12 Off-Road Vehicles (Wheeled and Tracked)

Günter H. Hohl

CONTENTS

12.1 Introduction 396
12.2 Classification 398
12.2.1 General 398
12.2.2 Wheeled Non-Armored Vehicles 399
12.2.3 Wheeled Armored Vehicles 399
12.2.4 Tracked Non-Armored Vehicles 400
12.2.5 Tracked Armored Vehicles 401
12.3 Vehicle–Terrain Interactions 401
12.3.1 Running Gear–Terrain Interaction 402
12.3.1.1 Terrain Specifications 402
12.3.1.2 Traction Forces 406
12.3.2 Vehicle Geometry—Terrain Interaction 415
12.3.2.1 Failure Modes 416
12.3.2.2 Vehicle Geometry 416
12.3.2.3 Overcoming Step-Formed Obstacles 418
12.4 Shocks and Vibrations 425
12.4.1 General 425
12.4.2 Human Response to Vibrations 427
12.4.2.1 General 427
12.4.2.2 Mechanical Models of the Human Body (Physical Effects) 427
12.4.3 Test Methods, Recommendations, and Standards 428
12.4.3.1 Absorbed Power 431
12.5 Technical Concepts 431
12.5.1 Vehicle Bodies 431
12.5.1.1 Frame 431
12.5.1.2 Hull 432
12.5.2 Power Train 434
12.5.2.1 General 434
12.5.2.2 Engine 435
12.5.2.3 Transmission 438
12.5.2.4 All-Wheel Drive and Differential Locks 440
12.5.2.5 Final Drive 443
12.5.2.6 Tires 443
12.5.2.7 Tracks 451
12.5.3 Suspension of Wheeled Terrain Vehicles 452
12.5.3.1 General 453
12.5.3.2 Non-Independent Suspension 454
12.5.3.3 Independent Suspension 456

12.5.3.4 Twisting Ability 460
12.5.3.5 Body Roll 465
12.5.3.6 Steering Ability and Wheel Control 467
12.5.3.7 Further Considerations and Summary 467
12.5.4 Suspensions of Tracked Vehicles 468
12.5.4.1 General 468
12.5.4.2 Torsion Bar Suspension 469
12.5.4.3 Coil Springs 470
12.5.4.4 Hydropneumatic Suspension 471
12.5.4.5 Dampers 471
12.6 Conclusion 472
References 473

12.1 INTRODUCTION

Automobiles are, at present, the most widely used means of transport on roads. When there are no roads, cargo and persons must be transported by off-road vehicles, which are designed for this purpose.

This chapter gives a general overview on models and theoretical and practical considerations for wheeled and tracked off-road vehicle systems. In addition to this, relevant elements of off-road vehicles and their influence on mobility are presented. Finally, the human factors and future developments are pointed out.

The *steam engine* was the first power source for vehicles on tracks and later on roads, then finally on terrain (Figure 12.1).

After the invention of the *internal combustion engines*, first according to the *Otto* and later to the *Diesel principle*, the steam engine lost its importance as a propulsion system for road vehicles as well as for off-road vehicles.

Off-road vehicles are used in large numbers, for example, in

- Agriculture
- Military operations
- Expeditions
- Rescue missions
- Construction
- Recreation activities

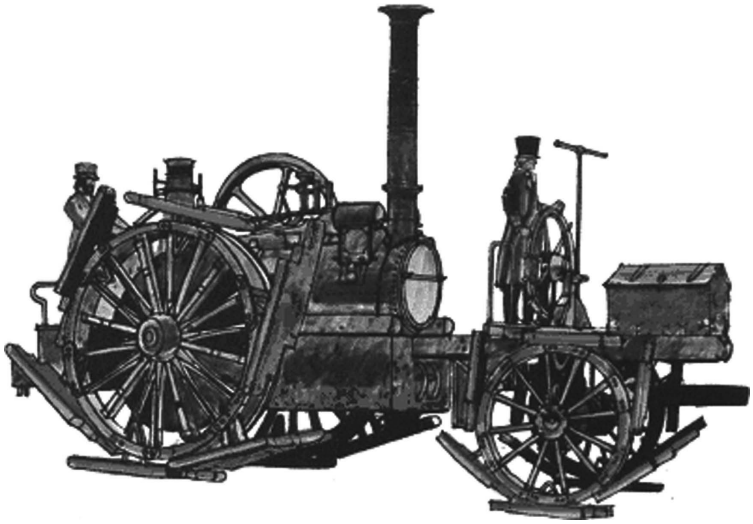


FIGURE 12.1 Burrell–Boydell tractor of 1857. (Reproduced from Ellis, C. and Bishop, D., *Military Transport of World War I*, Brandford Press Ltd., England, U.K.)

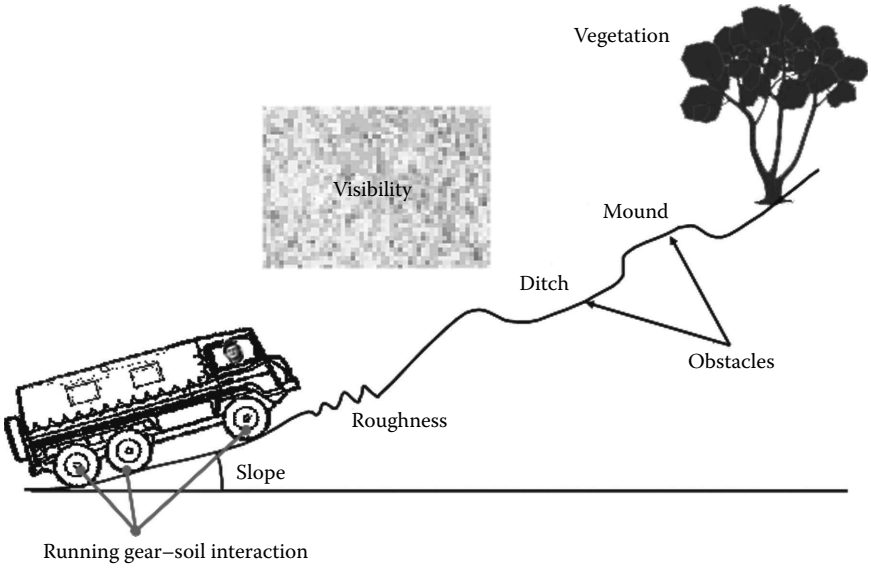


FIGURE 12.2 Vehicle in terrain interaction. (Reproduced from Hohl, G.H. and Corrieri, A., Basic considerations for the concepts of wheeled off-road vehicles, *FISITA World Automotive Congress*, Seoul, Korea, 2000 and based on an idea from *Vehicle Traction Mechanics*, Yong, R.N., Copyright, from Elsevier.)

Off-road vehicles share many components and design principles with commercial vehicles. However, in order to operate on terrain, these vehicles require different considerations, a new design or modifications of components.

The first considerations for a concept of off-road vehicles are to analyze the differences between road and terrain and the resulting influences.

The strength and deformation of terrain material must provide the necessary flotation and traction to keep a vehicle in constant travel and produce the required tractive force. Terrain characteristics can constitute other factors that could restrict vehicle speed or even cause total vehicle immobilization. These factors can be identified in Figure 12.2.

Influence of the terrain on the vehicle–terrain interaction [1]:

1. *Running gear–soil interaction*, "Running gear" *is the umbrella term for wheels a*nd tracks. The term "soil" is used in literature in connection with terrain vehicles to represent all types of transport surface. The determining factors for possible movement of off-road vehicles are the mechanical properties of the soil and the characteristics of the running gear system. The vehicle–terrain interaction is described in Section 12.3.
2. *Slope* can be anything up to a vertical wall, whereas roads rarely exceed 18%. Motion resistance increases because of the gravitational component of the vehicle along the slope. When traversing a slope, the vehicle could come into a dangerous overturning situation.
3. *Obstacles* include surface features, such as mounds, ditches, boulders, stumps, logs, dikes, and potholes, which are impediments to vehicles.
   There must be
   a. No interference between the obstacle and the bottom profile of the vehicle
   b. Enough traction and enough torque at the running gear
4. *Vegetation*, for example, grass-covered surfaces, causes additional resistance in vehicle motion and maneuvering around vegetation, like trees and bushes, requires additional traction forces and small steering radii.

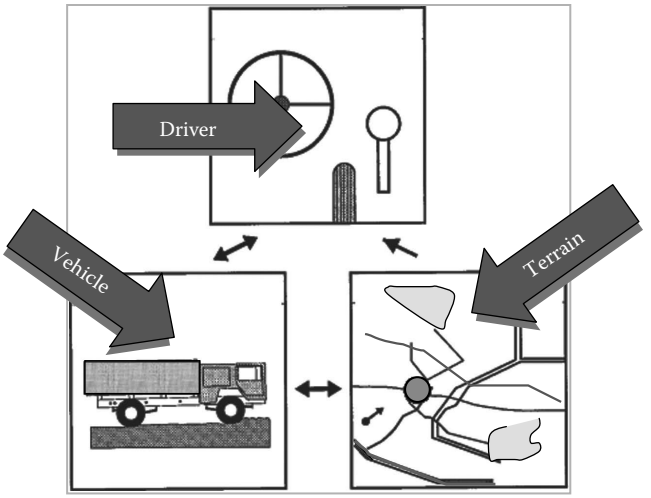


FIGURE 12.3 Driver–Terrain–Vehicle model. (Reproduced and modified from Ruff, K. and Jakobs, R., Interactive simulation of vehicle movement in terrain, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

5. *Visibility* conditions influence the driver's recognition distance. Visibility controls the maximum speed at which the driver may proceed.
6. *Roughness* can be defined as random ground surface irregularities, which are the source of shocks and vibrations to the vehicle body, to the running gear assembly and also to the crew (see Section 12.4).
7. *Water* in brooks, rivers, and lakes can be a hindrance for an off-road mission. It has to be distinguished between having contact with the ground with slow speed and floating, respectively. For the first situation, the vehicle needs a special design, so that water cannot damage any component of the vehicle. For amphibious abilities, even more complicated design concepts are necessary.

For basic considerations of a terrain vehicle, a simple model can be used, the so-called "Driver–Vehicle–Terrain" model. The three elements influence each other (Figure 12.3).

Another more detailed model is described in Ref. [2].

12.2 CLASSIFICATION

12.2.1 General

In technology, one always tries to form categories. These attempts have also been made for terrain and vehicles (see Sections 12.3 and 12.5).

Irrespective of unconventional off-road vehicles in niches, such as marsh screws, air cushion vehicles, and slides [3], conventional terrain vehicles can be divided with respect to their suspension systems into two main categories:

- Wheeled vehicles
- Tracked vehicles

Concerning the structure, there are two further categories:

- Non-armored (or soft-skinned) vehicles
- Armored vehicles

It is a fact that in soft ground conditions and for good ditch crossing ability, tracked vehicles are superior to wheeled vehicles. Generally, Main Battle Tanks (MBTs) are based on tracked suspension systems. Armored personnel carriers (APC) with infantry on board that fight together with MBT must, therefore, have the same level of mobility and are tracked vehicles. The disadvantage, in comparison with armored wheeled vehicles (AWV), is the higher price and the lower road speed.

For other purposes, as in reconnaissance and peace-keeping missions, wheeled armored vehicles have higher road speed and thus offset the lower off-road mobility. Due to the lower price, more vehicles can be introduced in military units.

12.2.2 Wheeled Non-Armored Vehicles

Wheeled non-armored (soft-skinned) vehicles are used for transport of persons and goods in poor terrain situations as well as on paved roads. The structure is a frame (see Section 12.5). These types of vehicles have a range from *S*port *U*tility *V*ehicles (SUV) to extreme heavy vehicles with a gross weight of up to 40 t and thus can be divided into light, medium, and heavy versions (Figure 12.4).

Two principal concepts in developing soft-skinned vehicles can be pointed out. The one group is specially designed for off-road purposes. The other category is based on civilian utility vehicles and modified in details for their terrain missions like greater fuel tanks, all-wheel drive, and differential locks.

12.2.3 Wheeled Armored Vehicles

Wheeled armored vehicles have an armored hull, which is the structure of the vehicle. The hull is the structure and can be regarded as an over dimensioned torsion-resistant tube frame (see Section 12.5.1)

They are equipped with small caliber arms and are used for transport of infantry men, reconnaissance, peace-keeping and police missions, and ambulance operations (Figure 12.5).

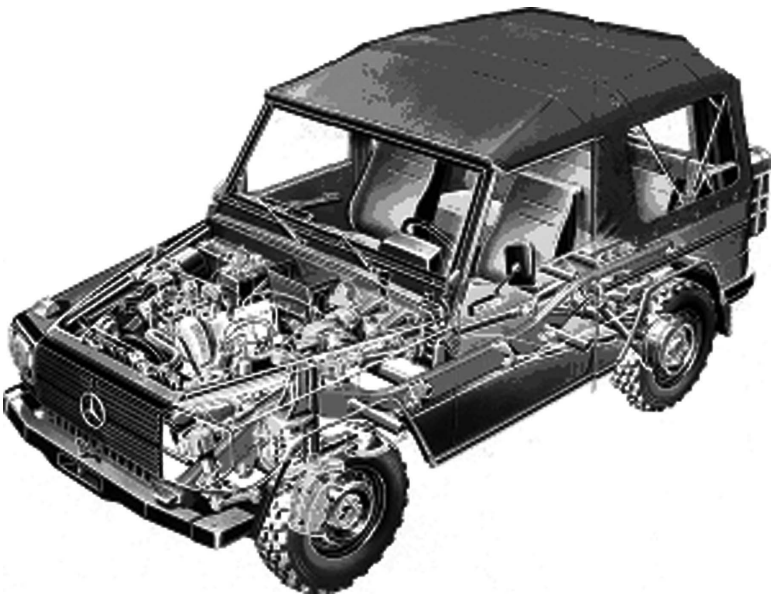


FIGURE 12.4 Puch Mercedes "G." (Reproduced from a leaflet from Steyr Daimler Puch A.G., Steyr, Austria.)

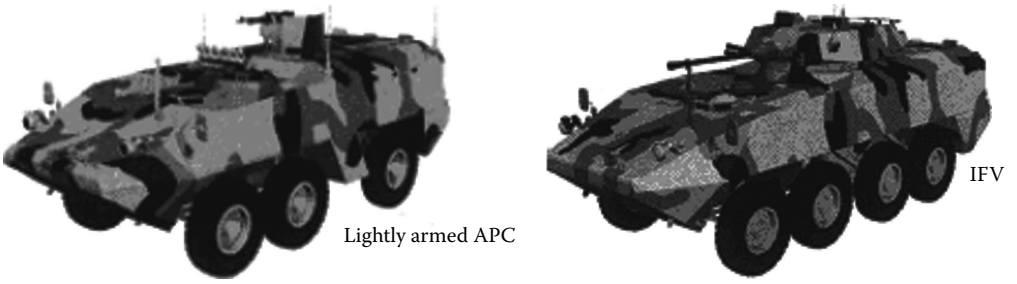


FIGURE 12.5 Family of wheeled armored vehicles. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

They can be divided into lightly armed APC with three axles, *I*nfantry *F*ighting *V*ehicles (IFV) with four axles, and *A*rmored *C*ombat *V*ehicles (ACV) with a five-axle concept, which can have almost the same tasks as MBTs.

12.2.4 Tracked Non-Armored Vehicles

Tracked non-armored vehicles are a niche. Contrary to armored versions, they have a kind of frame as the vehicle body. The oversnow vehicles in skiing areas belong to this one well-known category (Figure 12.6).

Another famous one and introduced in great number is the Swedish BV 206 with an articulated steering system (see Section 12.5.2.3.2). This vehicle can be used for transport of persons and/or goods, for ambulance purposes, and as a carrier for light weapons, radio, and light fire-fighting equipment (Figure 12.7).



FIGURE 12.6 Oversnow vehicle "Pisten Bully." (Reproduced from a leaflet from Kässborer A.G.)

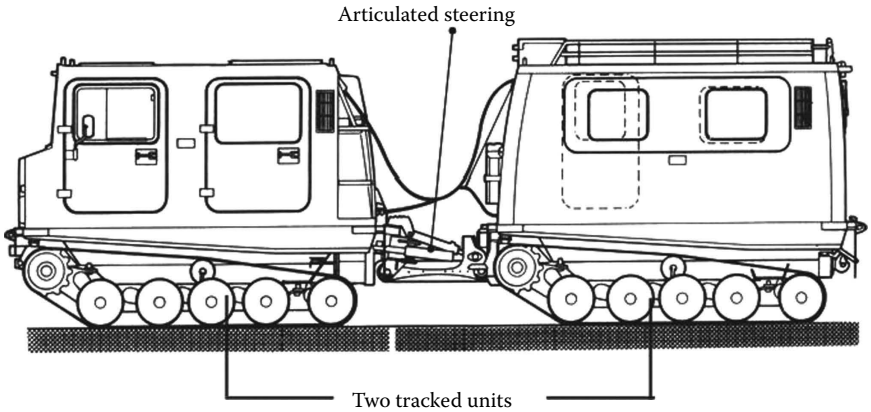


FIGURE 12.7 Swedish tracked vehicle "BV 206" with articulated steering. (Reproduced and modified from Wong, J.Y.: *Theory of Ground Vehicles*. 2001. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)



FIGURE 12.8 Modern MBT; "German *Leopard 2*." (Reproduced from Ford, R., *Panzer von 1916 bis heute*, Karl Müller Verlag, Köln, Germany.)

12.2.5 Tracked Armored Vehicles

Tracked armored vehicles can principally be divided into battle tank versions, with the engine compartment in the rear and in APC concepts with the engine compartment in the front. The main representatives of these species are the MBTs and APC.

Both types look similar: they are armored, have a turret with a weapon, and are tracked, but they have quite different missions and design principles (see Section 12.5.1.2; Figure 12.8).

History and technology of tracked armored vehicles are described in detail in Ref. [4].

Modern MBTs have suspensions with a high degree of energy absorption and high travel of the road wheels and a revolving turret. The engine is a diesel and in one block with the transmission–steering unit. (For more details about this concept, see Section 12.5.)

12.3 VEHICLE–TERRAIN INTERACTIONS

As mentioned in Section12.1, the off-road vehicle and the terrain influence each other.

This chapter describes the interactions in the contact area between the "running gear," which is the umbrella term for wheels and tracks, and the contacts between the terrain and parts of the vehicle.

12.3.1 Running Gear–Terrain Interaction

The first step in investigating the interaction is to define the properties of the soil.

Many theories and models are available for the estimation of the interaction, and each of these theories has advantages and disadvantages that are suitable for certain applications but unfortunately they are not related to each other.

12.3.1.1 Terrain Specifications

For the prediction of traction of an off-road vehicle, it is necessary to define the ground. There are three methods in use:

- Bekker values
- Cone pentrometer values
- Basic soil parameters

12.3.1.1.1 Bekker Values

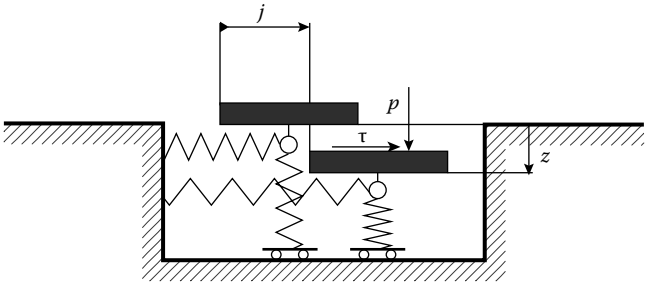
When observing movements of an off-road vehicle on soft ground, two things can be recognized:

- Sinkage
- Shear displacement

These two motions can be described by the plate model. A small piece has two movements in vertical (sinkage) and horizontal (shear) direction. The movements work against two symbolic springs. Contrary to springs in mechanical engineering, they have a plastic and a nonlinear characteristic. The stresses versus displacements in both directions can be regarded as one of the specifications of the soil properties (Figure 12.9).

The pressure–sinkage relationship can be characterized by the equation proposed by Bekker [2,5,6] and Wong [7,8]:

$$p = \left(\frac{k_c}{b} + k_\phi \right) z^n \tag{12.1}$$



j.....Shear displacement
z.....Sinkage
p.....Nominal normal pressure
τ.....Shear stress

FIGURE 12.9 Plate model. (Based on an idea in Krick, G., Wechselbeziehung zwischen starrem Rad, Luftreifen und nachgiebigem Boden, thesis, Munich, Germany, 1971; reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

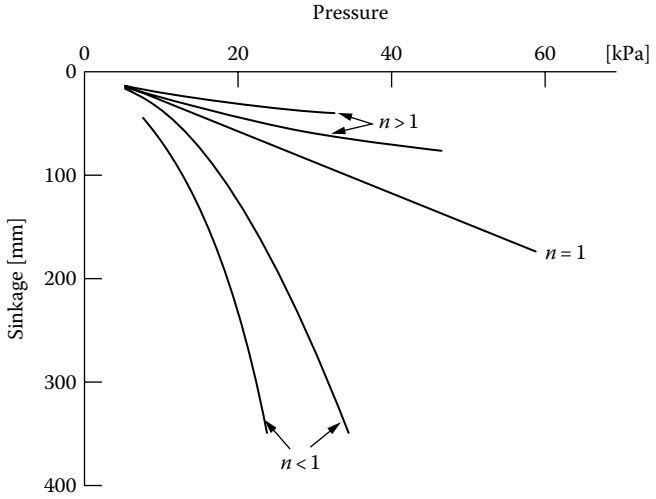


FIGURE 12.10 Nominal pressure–sinkage curves (parameter *n*). (Reproduced and modified from Wong, J.Y.: *Theory of Ground Vehicles*. 2001. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

where

p is the nominal pressure [kN/m^2]
b is the smaller width of a rectangular plate or radius of a circular plate [m]
z is the sinkage [m]
k_c is the cohesive modulus
k_ϕ is the frictional modulus
n is the pressure–sinkage parameter

Using the dimensions mentioned earlier, k_c has the dimension (kN/m^{n+1}) and k_ϕ has the dimension (kN/m^{n+2}), n is dimensionless. The soil parameters k_c, k_ϕ, n can be derived by a minimum of two tests with plates of different widths (Figure 12.10).

The pressure–sinkage relationship is strongly influenced by size and shape of the plate as well as by the velocity of the penetration [9].

The state of failure, which is the maximum shear stress (τ_{max}) in soils, is a function of the nominal pressure p and is expressed by the Mohr–Coulomb failure criterion:

$$\tau_{max} = c + p \tan \phi \tag{12.2}$$

where

p is the nominal pressure
c is the cohesion
ϕ is the angle of internal friction

The shear stress (τ)–shear displacement (j) relationship is a function of τ_{max}, which depends on the normal pressure. The τ–j relationship is described by an exponential function proposed by Janosi and Hanamoto:

$$\tau = \tau_{max}(1 - e^{-j/K}) = (c + p \tan \phi)(1 - e^{-j/K}) \tag{12.3}$$

The value K can be determined from the slope of the shear curve at the origin (Figure 12.11).

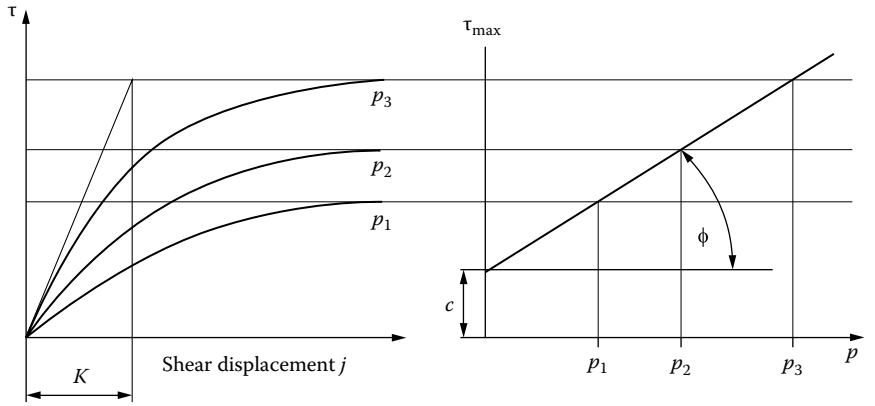


FIGURE 12.11 Shear stress–shear displacement relationship (Coulomb equation). (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria; Wong, J.Y.: *Theory of Ground Vehicles*. 2001. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

According to the Bekker theory, soil can be determined by the five values:

- Cohesive modulus, k_c
- Modulus of internal friction, $k_ø$
- Exponent, n
- Cohesion, c
- Angle of internal friction, ø

This method is also described in more detail in *Wong Mechanics*, Chapter 5.1, "Bevameter Techniques," as well as in Refs. [2,5–7]. A list of the Bekker values can be found in Ref. [7], Table 2.3.

The Bekker values can be used for predicting traction force–slip relationship and motion resistance of wheels and tracks (see *Wong Mechanics*, Chapters 5.2, 5.3, 6.1, and 6.2).

12.3.1.1.2 Cone Pentrometer

Cone pentrometers were developed for military purposes by the U.S. Army for the investigation of trafficability of soil toward the end of World War II. The cone pentrometer consists of a 30° circular cone with a base area of ½ in.^{2} (3.23 cm^2).

The measured values are expressed in the so-called *C*one *I*ndex (CI), which is the penetration force divided by the cone base area [1,3,7,10].

The necessary force is applied by pushing the cone slowly at a rate of about 30 mm/s through the soil. The force is measured by a deformable steel ring, which acts as a spring. The displacement is measured by dial gauge. With the spring characteristic, the penetration force can be determined (Figure 12.12 and Table 12.1).

Moisture content has a great influence on the CI values [11].

The Cone Pentrometer Method is used for empirical models like the *N*ATO *R*eference *M*obility *M*odel (NRMM), which is useful for the prediction of the mobility of military units. The advantage of the cone pentrometer method is that the terrain values can be easily measured. These values can also be used to investigate a "go or no go" situation and for predicting with an empirical method the drawbar coefficient μ_{20} at 20% slip for several soil types (see Section 12.3.1.2.2.4).

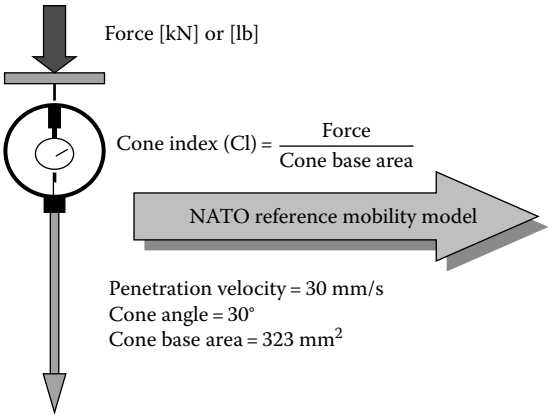


FIGURE 12.12 Cone pentrometer. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

TABLE 12.1
CI Values versus Soil Types

Soil Types	CI (N/cm²)
Coarse grained sand	200
Dry grassland	150
Sandy loam (medium dry)	120
Sandy loam (wet)	30–50
Ploughed field	20

Source: Reproduced and translated from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.

12.3.1.1.3 Basic Soil Properties

Soil is a multiphase medium that contains varying degrees and in various states of three distinct phases:

- Solid (inorganic and/or organic particles)
- Liquids (water and/or ice other fluids)
- Gas (air and/or other gases)

In natural soil the three phases are intimately mixed and shown schematically in Figure 12.13. The content of the three components can be expressed in percentage of volume or weight. The ratios between them define the basic physical soil properties.

The mostly used relationships are

- Porosity (n)
- Water content (w)
- Unit weight (γ)

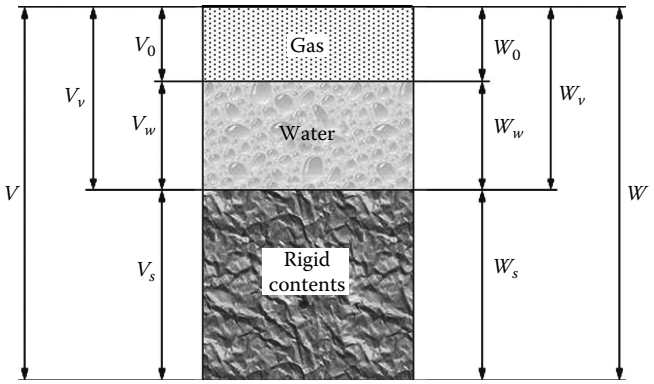


FIGURE 12.13 Idealized three-phase soil system. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria; *Vehicle Traction Mechanics*, Yong, R.N., Copyright, from Elsevier.; Karafiath, L.L. and Nowatzki, E.A., *Soil Mechanics for Off-Road Vehicle Engineering*, Trans Tech Publications, Aedermannsdorf, Switzerland.)

Porosity (n) in [%] is defined as the ratio of the voids $V_v = V_0 + V_w$ and the total volume:

$$n = \frac{V_0 + V_w}{V} 100 = \frac{V_v}{V} 100 \tag{12.4}$$

An easily measurable parameter that influences many of the soil characteristics is the *water content* (w) expressed as a percentage:

$$w = \frac{W_w}{W_s} 100 \tag{12.5}$$

The *unit weight* (γ) of the saturated soil–water system is defined by the total weight divided by the total volume:

$$\gamma = \frac{W_w + W_s}{V} = \frac{W}{V} \tag{12.6}$$

Unfortunately, it is not possible to define all characteristics of the soil usable for the interaction only by the physical values. There are also no generally valid relationships expressed in equations between the physical values, on the one hand, and the Bekker and CI values, on the other. But trends can be given: k_c, k_ϕ, c, and ϕ decrease with increasing water content w and increasing porosity n.

The CI increases with increasing unit weight γ and decreases with increasing water content w.

There were some attempts to find a relationship between the basic soil properties and Bekker values [1,5]. Hintze found a relationship for seasonal moisture and the CI values [11]. The properties of organic terrain are described in detail in Refs. [1,3,10].

12.3.1.2 Traction Forces

Before presenting calculation models for the prediction and methods for the measurement of traction forces of a terrain vehicle under certain ground conditions, it is necessary to make some definitions [1–3,5–7,12].

12.3.1.2.1 Definitions

The aim of an off-road ride is not only to move cross-country but also to produce traction for acceleration, to overcome obstacles and slopes, and to pull or push various types of trailers as agricultural implements, working and earthmoving machines, as well as military equipment like cannons. The force available at the drawbar is named drawbar pull (mostly abbreviated in literature with DP or F_d). The tractive effort F is the force directly acting at the running gear (wheel or track), which comes from the engine via all gear reductions and minus the internal friction. This force is due to the resisting forces ΣR not being fully available for pulling actions.

$$F_d = F - \sum R \tag{12.7}$$

ΣR includes all resistances like internal resistance of the running gear, the losses due to the interaction with the terrain, and the gradient and obstacle resistance. The tractive effort F is easy to determine when the torque–speed curve of the engine and the transmission characteristics are known.

DP is not constant, and it is a function of slip (i), which is defined as

$$i = \frac{v_{th} - v_a}{v_{th}} = 1 - \frac{v_a}{v_{th}} = 1 - \frac{v_a}{r\omega} \tag{12.8}$$

where

v_{th} is the theoretical speed ($r\omega$)
r is the radius
ω the angular velocity of the tire or of the sprocket
v_a is the actual speed
r is the radius of a wheel or sprocket

Slip i, which in some publications is also abbreviated with (s), can be expressed in decimal values between 0 and 1 or in percent values between 0% and 100%. Slip value 1, respectively, 100%, describes a rotating but immobile wheel or track.

The coefficient of traction μ_{tr}, which is a widely used parameter, is defined as the ratio of the DP F_d and the load on the driven wheels, respectively, the load on the track W_d:

$$\mu_{tr} = \frac{F_d}{W_d} = \frac{F - \sum R}{W_d} \tag{12.9}$$

The coefficient of rolling resistance μ_r is a parameter that is defined as the ratio of the sum of the rolling resistances and the load of the driven wheels, respectively, the load on the track W_d:

$$\mu_r = \frac{\sum R}{W_d} \tag{12.10}$$

Especially for working machines and tractors, the efficiency of the off-road vehicle in transforming the power produced by the engine and the power available at the drawbar is of economic interest. The tractive efficiency η_d is defined as the ratio of drawbar power P_d ($P_d = F_d v_{th}$) to the delivered power P by the engine:

$$\eta_d = \frac{P_d}{P} = \frac{F_d v_a}{P} = \frac{\left(F - \sum R\right) v_{th}(1-i)}{P} \tag{12.11}$$

This equation shows that the tractive efficiency η_d is high in cases of low slip values but in the case of 1, respectively, 100% slip, 0. Therefore, for a high tractive efficiency, the maximum of DP should be at low slip values.

12.3.1.2.2 Drawbar Pull–Slip Curves

12.3.1.2.2.1 Shape of the Curves The F_d (dB), respectively, the μ_{tr}–slip curves, have two principal shapes. In loose soils, progressive continuous characteristics up to 100% slip can be observed. Dense soils have a maximum in the range between 15% and 30% slip.

The magnitude and shape of a traction or traction force or coefficient of traction curve versus slip is the result of complex interactions between the running gear system and the terrain.

It is difficult to compare curves with completely different shapes as shown in Figure 12.14. In Ref. [13], four methods are described to evaluate the different curves (see Figure 12.15):

1. *Maximum* μ_{tr}—the maximum value or an average of a small amount of data (3% of the total data) surrounding the maximum.
2. *Average peak* μ—the optimum tractive performance, calculated by averaging the top 20% of the μ values.
3. μ *at 12% slip*—the value occurring when slip is 12%.
4. *SAE* μ—an average of the μ between a slip of 28% and 75%.

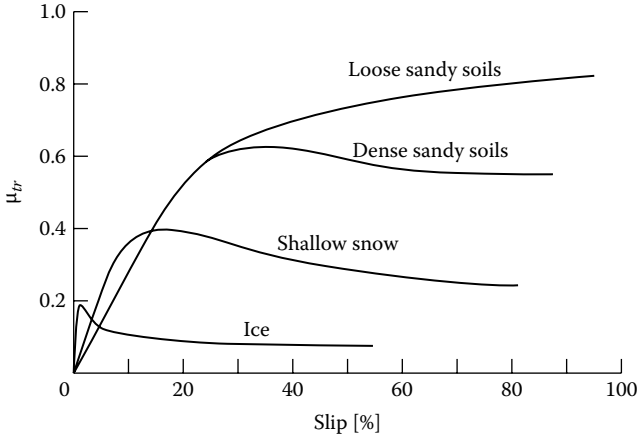


FIGURE 12.14 Generalized coefficient traction μ_{tr}–slip curves. (Reproduced from Shoop, S. et al., Winter traction using three different instrumented vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV, 1993.)

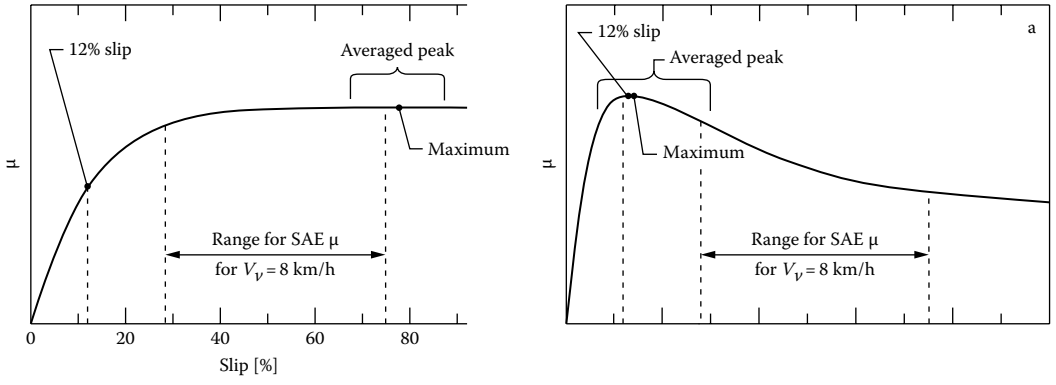


FIGURE 12.15 Calculation methods for evaluation of μ–slip curves. (Reproduced from Shoop, S. et al., Winter traction using three different instrumented vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV, 1993.)

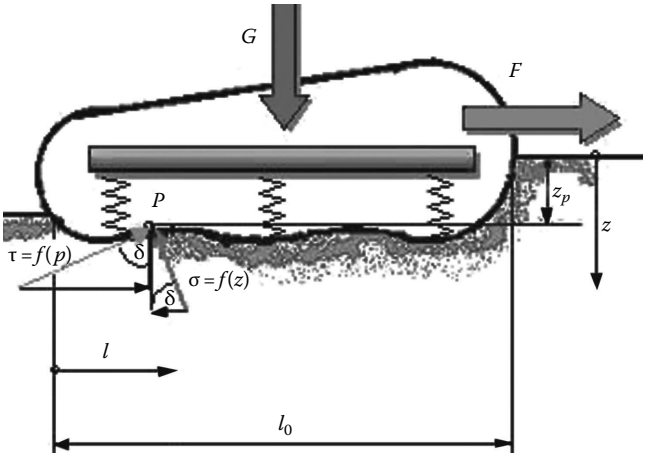


FIGURE 12.16 Principle of the SIM. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

12.3.1.2.2.2 Stress Integrating Method The classical *S*tress *I*ntegrating *M*ethod (SIM) initiated by Bekker [2,5,6] and developed by Wong [7,8], who approaches the phenomena of the running gear—soil interaction in more detail—is based on the Bekker terrain measurements.

To evaluate the relationship between the DP and the slip of a running gear system (wheel or track), the normal and shear stress distributions, depending on slip along the interface, have to be determined first. The vertical components of both stress types represent the load (*G*) and the horizontal components are equal to the tractive force (Figure 12.16).

This method can also be used for rigid wheels [1,7,8] as well as for tracked vehicles [7,8] and is described in detail in Refs. [7,8].

The flexibility of a pneumatic tire can be described with different contact shapes [14–18].

Some other scientists also developed models for the prediction of tractive force in soft soil and organic surface based on other considerations [14,19].

12.3.1.2.2.3 Finite Element Method The Finite Element Method (FEM) is widely used for stress determination in complex elements in mechanical engineering. In recent years, many attempts have been made to apply this successful method to the analysis for the running gear–terrain interaction. Professor Schmid and his assistants were very successful in this method [16,20–23].

The FEM investigates the phenomena of wheel–soil interaction in detail by defining a rolling contact between an elastic tire and soft soil (Figure 12.17).

The terrain is represented by a system of elements in triangular or rectangular form with specific constitutive relationships and interconnections at nodes to form a mesh. In the soil model, the soil has elastoplastic properties and the strength limits in a three-dimensional state are defined according to the Drucker–Prager theory. Based on this consideration, the soil FEM can be applied for frictional and cohesive soil. The tire is considered as piecewise linear elastic material for tread carcass and rim in the form of three concentric rings.

Similar to the FEM is the Distinct (Discrete) Element Method. The term discrete element method (DEM) is a family of numerical methods for computing the motion of a large number of particles like molecules or grains of sand. This method used for terramechanics is described in more detail in Refs. [24,25].

12.3.1.2.2.4 Empirical Method With this method, the tractive performance of a running gear is based on an empirical model. The dimensionless terms, the so-called numerics are related to the drawbar coefficient and the drawbar efficiency both at 20% slip. The maximum of DP is usually in this range. Therefore, these two dimensionless values lead to the index 20 (μ_{20} and η_{20}) [1,3,7,26,27].

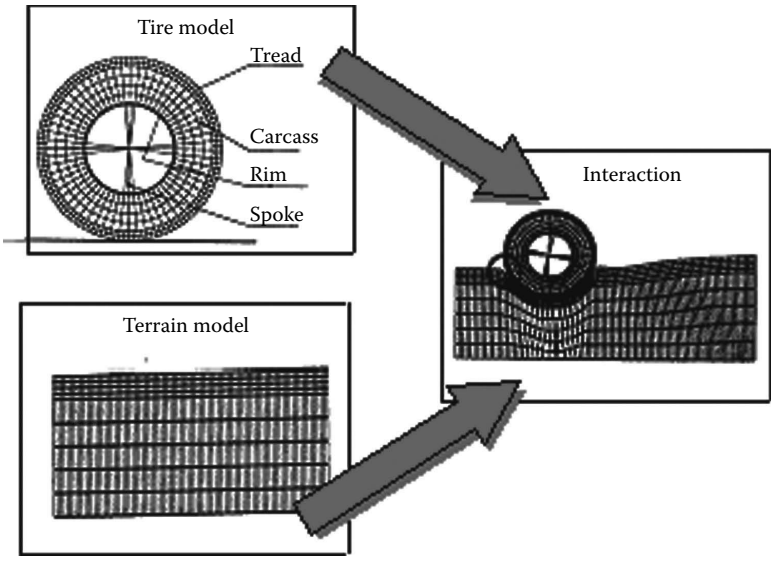


FIGURE 12.17 Principle of the FEM. (Reproduced and modified from Schmid, I., Interaction of vehicle and terrain, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

The "numerics" include the terrain data with the CI (in this model, the CI is abbreviated with C) and the tire data (diameter d, width b, section height h, deflection δ, and wheel load W). Depending on the soil type, three "numerics" are defined (Figures 12.18 and 12.19):

Clay–tire numeric N_c

$$N_c = \frac{C(bd)(d/h)^{1/2}}{W} \cdot \frac{1}{1+(b/2d)} \tag{12.12}$$

Sand–tire numeric N_s

$$N_s = \frac{G(bd)^{3/2}}{W} \cdot \frac{\delta}{h} \tag{12.13}$$

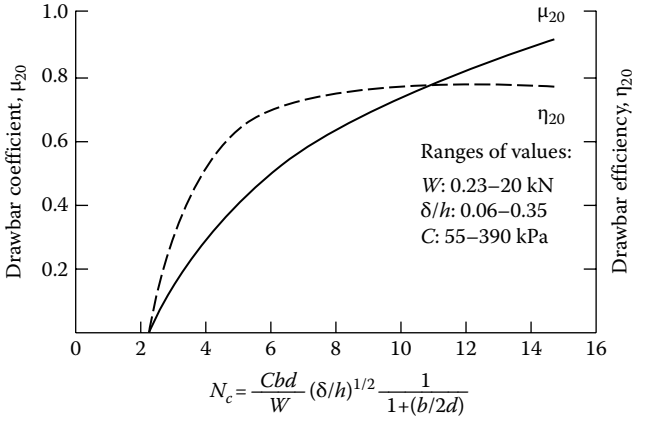


FIGURE 12.18 Empirical relationship between drawbar coefficient and drawbar efficiency at 20% slip for the clay–tire numeric N_c (left). (Reproduced from from *Vehicle Traction Mechanics*, Yong, R.N., Copyright, from Elsevier.)

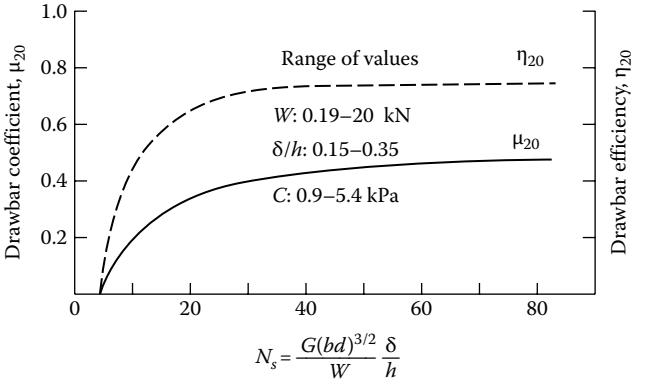


FIGURE 12.19 Empirical relationship between drawbar coefficient and drawbar efficiency at 20% slip for the frictional sand numeric N_s (right); G is the sand penetration resistance gradient. (Reproduced from *Vehicle Traction Mechanics*, Yong, R.N., Copyright, from Elsevier.)

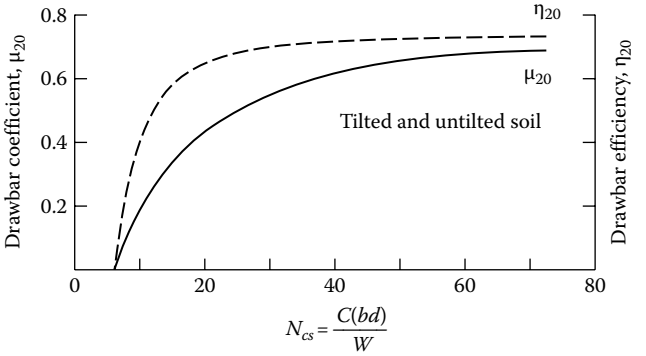


FIGURE 12.20 Empirical relationship between drawbar coefficient and drawbar efficiency at 20% slip for the cohesive frictional sand numeric N_{cs}. (Reproduced from *Vehicle Traction Mechanics*, Yong, R.N., Copyright, from Elsevier.)

Cohesive and frictional sand–tire numeric N_{cs} (Figure 12.20)

$$N_{cs} = \frac{C(bd)}{W} \tag{12.14}$$

12.3.1.2.2.5 Practical Method To obtain realistic DP–slip curves, it is necessary to test the running gear under in situ situations. There are several ways available:

Test of a single wheel in a soil bin under laboratory conditions

- Single wheel tester on a natural surface
- Test rig pulling a braking (parent) vehicle
- Test rig pushing a braking vehicle
- Test of an overall vehicle

The tasks of a single wheel test are to measure only the interaction of a tire with the soil, without influences from the other components of the vehicle. In soil bins, the soil has laboratory conditions with constant and defined soil properties (CI, water content, porosity etc.) (Figure 12.21).

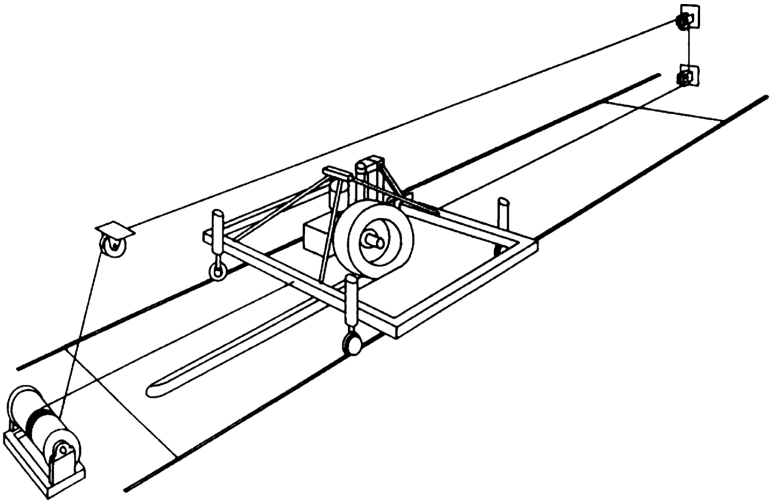


FIGURE 12.21 Principle of a soil bin. (Reproduced from Schwanghart, H. and Zweier, A., Influence of tread on drawbar pull of tires for small off-road vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV.)

Soil Bin Tests

The test tire is mounted on a single wheel test rig running on rails and towed by a cable winch. The test wheel is driven by a hydraulic or an electric motor. The wheel load is applied by a constant ballast weight. The wheel rotates with a constant speed. By varying the forward speed of the test rig, any slip between 0% and 100% can be set. Measuring wheel torque, DP, load, sinkage, wheel speed (v_{th}) and wagon speed (v_a), and tractive force, respectively, DP–slip curves can be determined (Figure 12.22).

Outdoor Single Wheel Tester

Contrary to the soil bin tests, the single wheel testers described in the following work on a horizontal terrain under natural soil conditions, which are prepared to a certain degree.

There are two types of single wheel testers:

- Test device with a towed braking (parent) vehicle
- Test device in a frame with a pushed braking vehicle

In both cases, the tester consists of two vehicle bodies. The driving part is the test tire and the driven part is pushed or pulled and has a braking function. The force in between is the DP. The motion of the braking vehicle is without slip and thus represents the actual speed.

The single wheel tester designed by Royal Armaments Research & Development Establishment (RARDE) is shown in Figure 12.23. The force P between the two vehicle bodies is the DP and due to the measurement of torque, the tractive force can be determined.

The theoretical speed v_{th} is measured at the hub of the test wheel and the actual speed v_a at the hub of the parent vehicle.

The parent vehicle is pushed by the hydraulically driven test wheel. With this test equipment, not only the tractive force but also the lateral force can be measured (Figures 12.24 and 12.25).

Overall Vehicle Tests

The advantage of an overall vehicle test is that the soil conditions are realistic and the procedures for testing and measuring are easier. On the other hand, however, the soil conditions are not as equal in property as in a soil bin, the speed is not steady, and, therefore, the acceleration forces are included.

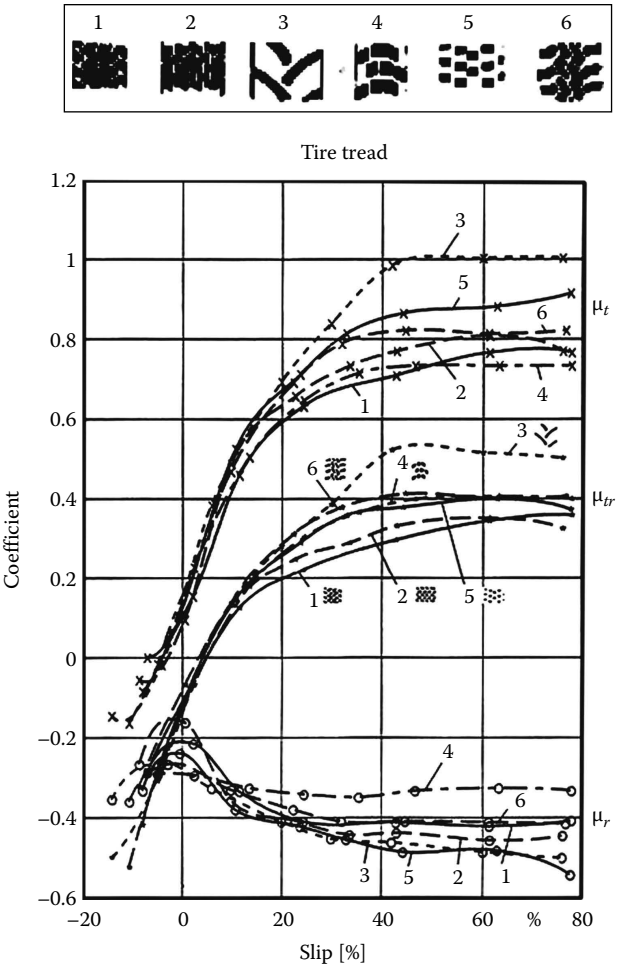


FIGURE 12.22 Test results of a soil bin test with six test tires; the numbers represent different tires. (Reproduced and modified from Schwanghart, H. and Zweier, A., Influence of tread on drawbar pull of tires for small off-road vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV.)

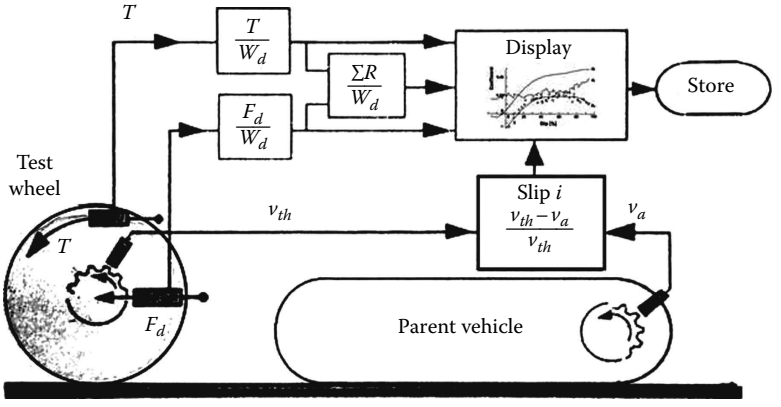


FIGURE 12.23 RARDE single wheel tester; test results. (Reproduced and modified from Maclaurin, E.B., A rig for testing the soft soil performance of tracked systems, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)



FIGURE 12.24 University of Stuttgart–Hohenheim single wheel tester; test arrangement. (Reproduced from a picture from University of Stuttgart–Hohenheim, Stuttgart, Germany.)

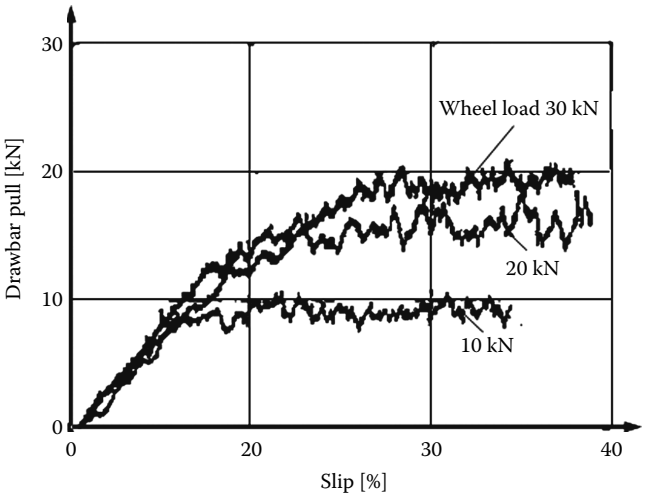


FIGURE 12.25 University of Stuttgart–Hohenheim; single wheel tester; test results. (Reproduced and modified from Kutzbach, H.D., Investigation on tractor tyres test stand and results, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

The test vehicles are equipped with test tires towing a braking (parent) vehicle. The DP is measured by a force transducer attached to the rope between the two vehicles.

The principal way of measuring DP, theoretical and actual speed, load, and in some cases torque is similar for the test methods mentioned earlier.

Load is applied by weights. For all types of single wheel testers, it is not a problem to measure the theoretical speed as it is proportional to the revolutions of the wheels or the sprocket. It is more difficult to measure the actual speed in field tests with a complete vehicle. One widespread method is to use a towed wheel also named "Peiseler" or "fifth wheel." The disadvantage is that such a wheel can temporarily lose ground contact and that the wheel itself can be subject to slip.

Another method, which was used by the author, is a noncontact sensor system. The contactless operation of the measuring device fully preserves the maneuverability of the vehicle in all directions. DP is measured by a dynamometer hook fixed to a rigid point. In the case of single wheel testers and tests with a complete vehicle, the dynamometer hook is attached between the test vehicle and the braking vehicle.

In the tests mentioned in Figure 12.26, the actual speed was measured by a contactless measuring device and for the determination of the DP, the towing forces and the accelerations of the vehicle mass were taken into consideration. Knowing the DP according to the soil properties and the

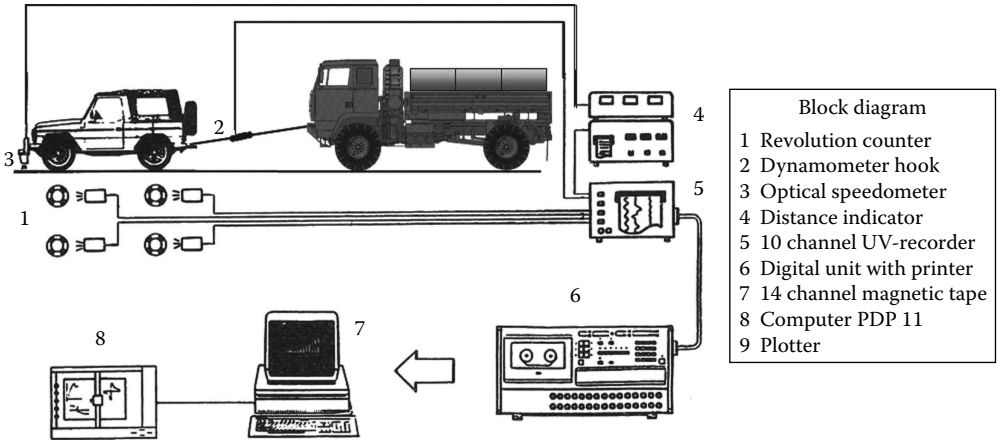


FIGURE 12.26 Austrian Agency of Military Technology; test arrangement. (Reproduced and modified from Hohl, G., A method to determine coefficient of traction-slip relationship of tyres in "in-situ" tests, *9th International ISTVS Conference*, Barcelona, Spain.)

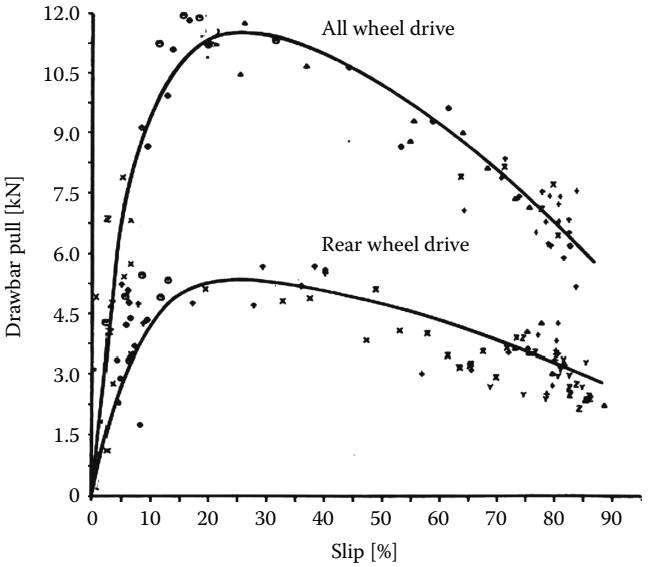


FIGURE 12.27 Austrian Agency of Military Technology; test results. (Reproduced and modified from Hohl, G., A method to determine coefficient of traction-slip relationship of tyres in "in-situ" tests, *9th International ISTVS Conference*, Barcelona, Spain.)

different contact shapes [7,8,16–18], the maximum speed and the acceleration can be determined. For the prediction of acceleration, the translatory and rotatory masses have to be taken into consideration. There are two ways to combine these two masses:

- To put the translatory masses to the rotatory ones [28,29]
- To put the rotatory masses to the translatory ones [7,30] (Figure 12.27)

12.3.2 Vehicle Geometry—Terrain Interaction

The "waviness" of the ground obviously may degenerate into small irregularities, which cause vibrations (see Section 12.4) and deep cut forms that may stop the vehicle motion entirely. These geometrical obstacles are described in Section 12.3.2.1.

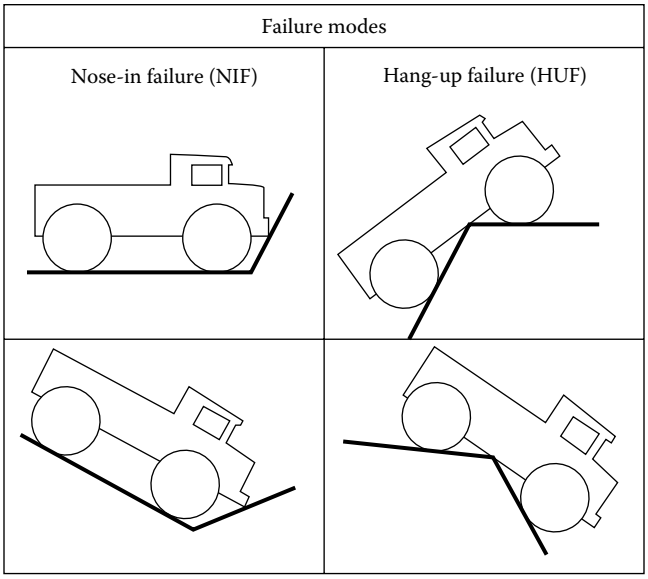


FIGURE 12.28 Vehicle–terrain failure modes. (Reproduced from Bekker, M.G., *Introduction to Terrain-Vehicle Systems*, University of Michigan Press, Ann Arbor, MI, 1969.)

12.3.2.1 Failure Modes

Due to the unevenness of the terrain, parts of the vehicle can come into contact with the ground and can be immobilized independently from the physical conditions of the surface. There is a great variety in the form and sizes of natural terrain obstacles that may immobilize a terrain vehicle. According to Bekker [2], the mechanics of interaction of these two geometries must be defined in any characterization of terrain roughness from the viewpoint of

- Obstacle crossing
- Vehicle controllability
- Ride comfort (see Section 12.4; Figure 12.28)

With the two principal failure modes defined by Bekker [2], nearly all geometrical vehicle terrain conflicts can be described:

- Nose-in failure (NIF)
- Hang-up failure (HUF)

An NIF happens when the front or the rear end interferes with the obstacle.

In case of an HUF, the bottom of a vehicle comes into contact with an obstacle. Both types occur on convex or concave obstacle forms.

12.3.2.2 Vehicle Geometry

To avoid an NIF, it is important that the terrain vehicle has high angles of approach and of departure.

The *angle of approach* is defined as the maximum angle that an obstacle can overcome without striking the obstacle and damaging the vehicle.

The *angle of departure* is the maximum angle at which the vehicle can exit an obstacle without striking the obstacle and damaging the vehicle.

For a vehicle with good off-road abilities, both values should be in the range of 30°–40° or more (Figure 12.29).

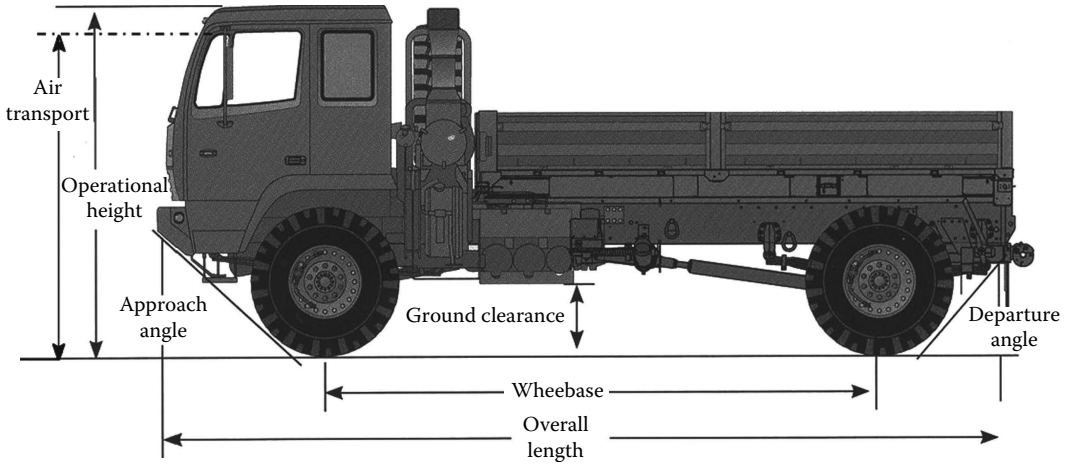


FIGURE 12.29 Definition of departure angle and approach angle. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

An HUF situation occurs when a vehicle moves over a wedge-formed obstacle. The shape of the ridge is defined by the breakover angle, which is the maximum angle a vehicle can traverse an obstacle without striking or hanging up between the axles. To find a geometrical relationship between the hung-up diameter D_r and the breakover angle β and the wheel diameter D, the axle distance l, and the ground clearance h_g, it is easier to move the wedge underneath the vehicle bottom.

The geometry of the wedge can be defined as ground clearance diameter D_r or as breakover angle β. There is a complicated relationship between both and the vehicle dimensions, axle distance l, and wheel diameter D. The breakover angle should be more than 20°:

$$D_r = \frac{2l^2 D(\cos\beta - \cos^2\beta)}{4l^2 \sin\beta - D^2(\cos^2\beta - 2\cos\beta + 1)} + \sqrt{\left[\frac{2l^2(\cos\beta - \cos^2\beta)}{4l^2 \sin^2\beta - D^2(\cos^2\beta - 2\cos\beta + 1)}\right]^2} + \frac{4l^4}{4l^2 \sin^2\beta - D^2(\cos^2\beta - 2\cos\beta + 1)} - D \tag{12.15}$$

In accordance with Figure 12.30 the vehicle will fail if

$$h_g \le 0.5\left[(D + D_r) - \sqrt{(D + D_r)^2 - l^2}\right] \tag{12.16}$$

The relationships in the equations earlier show that a high clearance diameter D_r, respectively, a low wedge angle β and high wheel diameter D and a short axle distance l are positive for overcoming a wedge-formed obstacle.

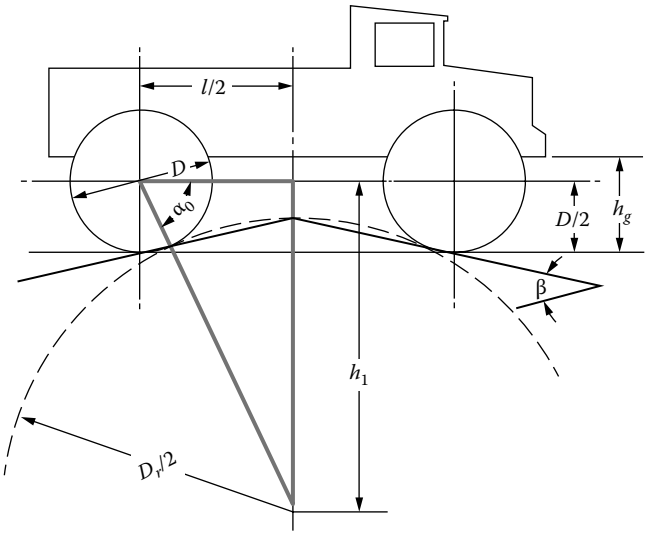


FIGURE 12.30 Longitudinal clearance diameter D_r and breakover angle β. (Reproduced from Bekker, M.G., *Introduction to Terrain-Vehicle Systems*, University of Michigan Press, Ann Arbor, MI, 1969.)

Transverse ground clearance is measured between the wheels on one axle. It can be defined by the diameter D_{rt} of the circle that touches the inner side of the tires and the lowest hanging point under a vehicle (usually a differential or the exhaust system).

Another definition is the zenith of the segment of the circle, which touches the wheels and the lowest point of the vehicle under the body (Figure 12.31).

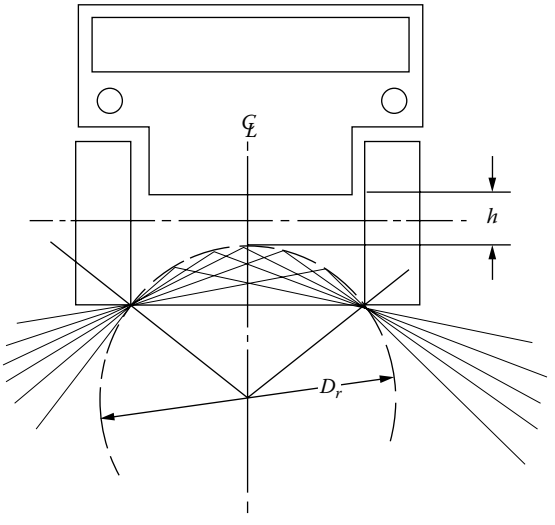


FIGURE 12.31 Transverse clearance diameter D_r. (Reproduced from Bekker, M.G., *Introduction to Terrain-Vehicle Systems*, University of Michigan Press, Ann Arbor, MI, 1969.)

- Vertical walls (A) and (B)
- Horizontal cavities (C)

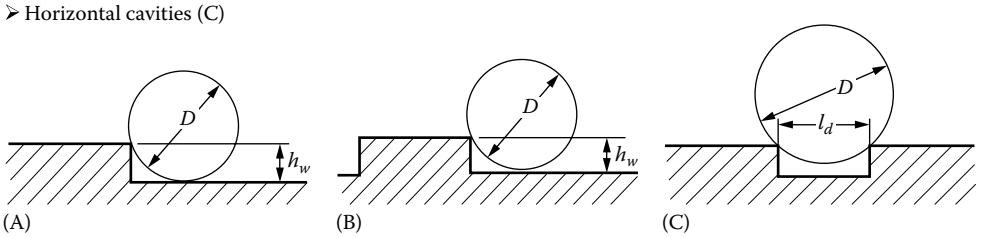


FIGURE 12.32 Vertical walls (A) Plateau type wall, (B) Bump type wall, and (C) Horizontal cavity. (Reproduced from N.N., *Engineering Design Handbook*, Automotive Series, Automotive Suspensions, AMCP 706–356, Headquarters, U.S. Army Materiel Command, Huntsville, AL.)

12.3.2.3 Overcoming Step-Formed Obstacles

Bekker [6] defined two elementary types of obstacles related to the geometry of terrain surface (Figure 12.32):

The described model with rigid wheels and no spring reactions is two-dimensional and the terrain and the vehicle are described by the following data:

Terrain	*Vehicle*
• Height of the step (h_w)	• Tire diameter (D)
• Coefficient of traction (μ_a)	• Wheel base (s)
• Coefficient of rolling resistance (μ_r)	• Position of C.G. (s_1, h_0)

The geometric situation when the front wheel encounters the wall is different from that when the rear wheel overcomes the wall.

12.3.2.3.1 Two-Axle All-Wheel Drive

The main principle of investigating climbing ability over a step is similar for all situations and is demonstrated in Figure 12.33 for a plateau-type wall by means of a two-axle all-wheel drive vehicle just entering the step.

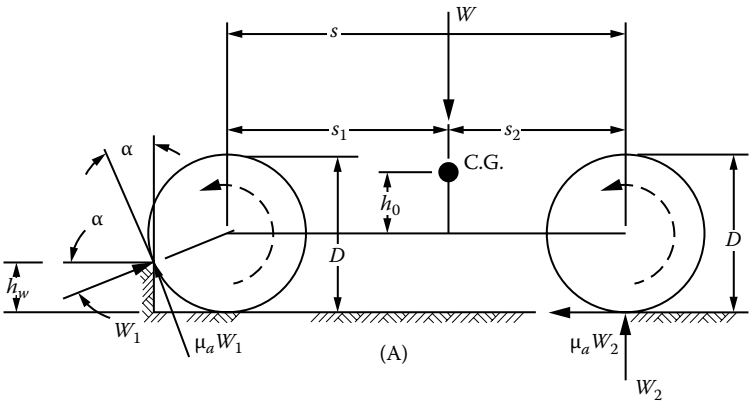


FIGURE 12.33 Two-axle vehicle 4 × 4 encountering a step in situation A. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria; Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956; N.N., *Engineering Design Handbook*, Automotive Series, Automotive Suspensions, AMCP 706–356, Headquarters, U.S. Army Materiel Command, Huntsville, AL; Hohl, G.H. and Peschak, B., A model to predict the overcoming of geometrically defined obstacles by a two-axle wheeled vehicle, *13th International ISTVS Conference*, TU München, Munich, Germany, 1999.)

The sum of all horizontal and vertical forces and the sum of the momentum around the axis of the front wheel *I* are zero. This way of solving the equations is hereby described in Figure 12.33 and shows situation A when the front wheel hits edge A of the step. The equilibrium of forces and momentums yield the following equations:

$$\begin{aligned} W_1 \cos\alpha - \mu_a W_1 \sin\alpha - \mu_a W_2 &= 0 \\ W_1 \sin\alpha + W_2 + \mu_a W_1 \cos\alpha - W &= 0 \\ -\mu_a W_1 \frac{D}{2} + W_2 s - W s_1 - \mu_a W_2 \frac{D}{2} &= 0 \end{aligned} \tag{12.17}$$

Solving the three equations (Equations 12.17) with three unknown values yields to the following implicit relationship:

$$\left[\frac{1}{\mu_a} - \frac{1+\mu_a^2}{\mu_a}\left(\frac{s_1}{s}\right) - \frac{D}{2s}\right]\cos\alpha - \left[1 - \frac{\mu_a D}{2s}\right]\sin\alpha - \frac{\mu_a D}{2s} = 0 \tag{12.18}$$

$$\sin\alpha = 1 - \frac{2h_w}{D} \tag{12.19}$$

$$\cos\alpha = \frac{2\sqrt{Dh_w - h_w^2}}{D} \tag{12.20}$$

$$T_1 = \left[\frac{1}{\mu_a} - \frac{1+\mu_a^2}{\mu_a}\left(\frac{s_1}{s}\right) - \frac{D}{2s}\right] \tag{12.21}$$

$$T_2 = \left(1 - \frac{\mu_a D}{2s}\right) \tag{12.22}$$

Using the auxiliary terms T_1 and T_2 (Equations 12.21 and 12.22) and the geometrical relationships (Equations 12.19 and 12.20) leads to Equation 12.23, where the vehicle data D, s, s_1, as well as the terrain data h_w and μ_a are linked together:

$$h_w = \frac{D}{2} \times \frac{T_1^2 + T_2 - \sqrt{\left(T_1^2 + T_2\right)^2 - \left(T_1^2 + T_2^2\right)}}{T_1^2 + T_2^2} \tag{12.23}$$

12.3.2.3.2 Rear-Wheel Drive

For rear-wheel drive, the traction coefficient μ_a for the front wheel has to be replaced by the coefficient of rolling resistance—μ_r in Equation 12.18.

This yields to the implicit relationship:

$$\frac{\mu_r D}{2s} = \left[\frac{S_1}{S}\left(\frac{\mu_a + \mu_r}{\mu_a}\right) - \frac{\mu_r}{\mu_a} + \frac{\mu_r D}{2s}\right]\sin\alpha - \left[\frac{1}{\mu_a} - \frac{1 - \mu_a\mu_r}{\mu_a}\left(\frac{S_1}{S}\right) - \frac{D}{2s}\right]\cos\alpha \tag{12.24}$$

h_w, respectively, h_w/D, can be solved in an iterative way by using the geometrical relationships in Equations 12.21 and 12.22. Assuming that μ_r is very small and thus can be set to zero, an explicit solution of this equation can be obtained between h_w/D and the vehicle and terrain parameters. This would be correct for hard ground.

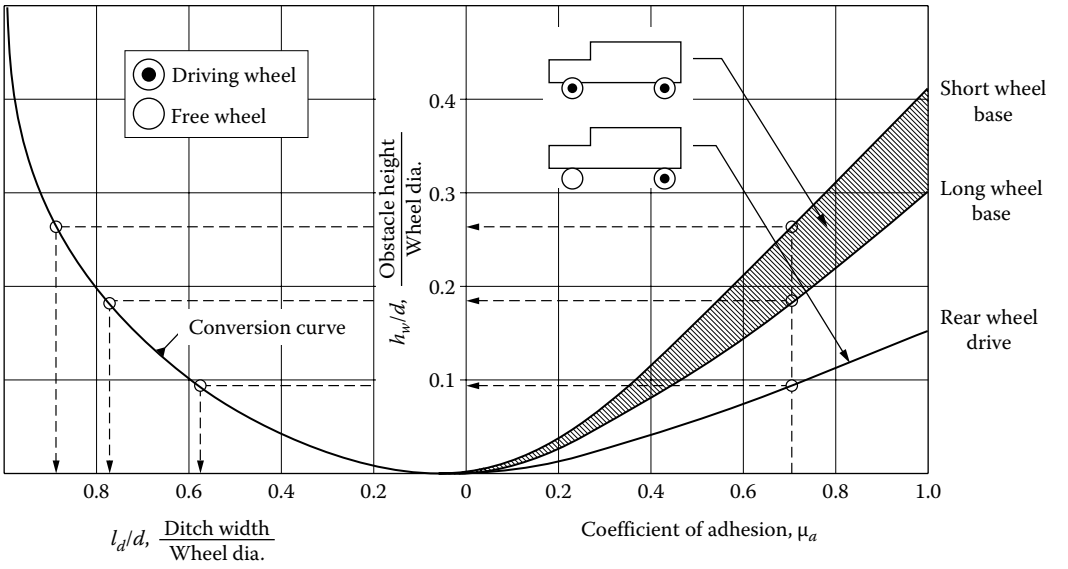


FIGURE 12.34 Step overcoming with parameters μ_a in situation A and ditch overcoming with l_d/d. (Reproduced and modified from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956.)

It is very useful to compare the various vehicle parameters by using the h_w/D relationship.

$$\left(\frac{h_w}{D}\right) = \frac{1}{2}\left\{1 - \frac{1}{\sqrt{1+\mu_a^2\left[(s_1/s)/1-(s_1/s)-(\mu_a D/2s)\right]}}\right\} \tag{12.25}$$

Crossing a ditch with the width l_d and a wheel diameter D leads to a similar situation as described in Figure 12.34 and Equation 12.26. The geometrical relationships between l_d/D and h_w/D are

$$\frac{l_d}{D} = 2\sqrt{\frac{h_w}{D} - \left(\frac{h_w}{D}\right)^2} \tag{12.26}$$

This leads to a relationship between μ_a, respectively, l_d/D to h_w/D.

The situation B in Figure 12.35 shows the rear wheel of an all-wheel drive vehicle overcoming a step. The geometrical situation in combination with equilibrium of forces and momentums yields to the following equations:

$$\begin{gathered} W_1 + W_2\sin\alpha + \mu_a W_2\cos\alpha - W = 0 \\ -\mu_a W_1 + W_2\cos\alpha - \mu_a W_2\sin\alpha = 0 \\ -W_1 s\cos\beta - \mu_a W_1\left(\frac{D}{2} - h_w\right) - \mu_a W_2\left(\frac{D}{2}\right) + W[(s-s_2)\cos\beta - h_0\sin\beta)] = 0 \end{gathered} \tag{12.27}$$

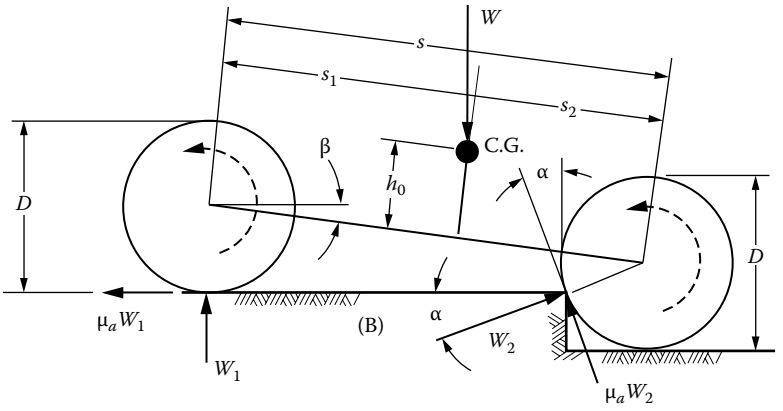


FIGURE 12.35 Two-axle vehicle 4 × 4 encountering a step in situation B. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956; N.N., *Engineering Design Handbook*, Automotive Series, Automotive Suspensions, AMCP 706–356, Headquarters, U.S. Army Materiel Command, Huntsville, AL; Hohl, G.H. and Peschak, B., A model to predict the overcoming of geometrically defined obstacles by a two-axle wheeled vehicle, *13th International ISTVS Conference*, TU München, Munich, Germany, 1999.)

Solving the Equations 12.27 yields to the following implicit relationship:

$$\left[\left(\cos\beta-\mu_a\sin\beta\right)+\frac{\mu_a D}{2s}\right]\sin\alpha-\left\{\left(\frac{1+\mu_a^2}{\mu_a}\left(\frac{s_1}{s}\right)-\mu_a\right)\cos\beta\right\}-$$

$$-\left\{\left[\frac{1+\mu_a^2}{\mu_a}\left(\frac{h_0}{s}\right)+1\right]\sin\beta-\frac{D}{2s}\right\}\cos\alpha=\frac{1}{2}\mu_a\frac{D}{s}\left[(s-s_1)\cos\beta+h_0\sin\beta\right] \tag{12.28}$$

$\sin\alpha$ and $\cos\alpha$ are as before in Equations 12.19 and 12.20.

With the following relationships of Equations 12.29 and 12.30, the maximum obstacle height h_w can be determined in an iterative way:

$$\sin\beta=\frac{h_w}{s} \tag{12.29}$$

$$\cos\beta=\sqrt{1-\left(\frac{h_w}{s}\right)^2} \tag{12.30}$$

The analysis of Equations 12.28 in combination Equations 12.19, 12.29, and 12.30, shows that the effect of s_1/s is opposite when the front wheels or the rear wheels of the four-wheel-driven vehicle negotiates an obstacle. A nose-heavy vehicle with a long wheel basis will perform it better with the rear wheels than when crossing it with the front wheels. Large s/D values improve the performance of the rear, regardless of the weight distribution.

Figure 12.36 indicates the effect of weight distribution s_1/s and the height of the C.G. on an obstacle climbing ability of the front wheels (situation A) and the rear wheels (situation B). For situation A, the C.G. near to the rear wheels is better; in situation B, a nose-heavy vehicle is better.

As the location of the C.G. cannot be changed while overcoming a step, the best performance is when the C.G. is in the middle of the two axles.

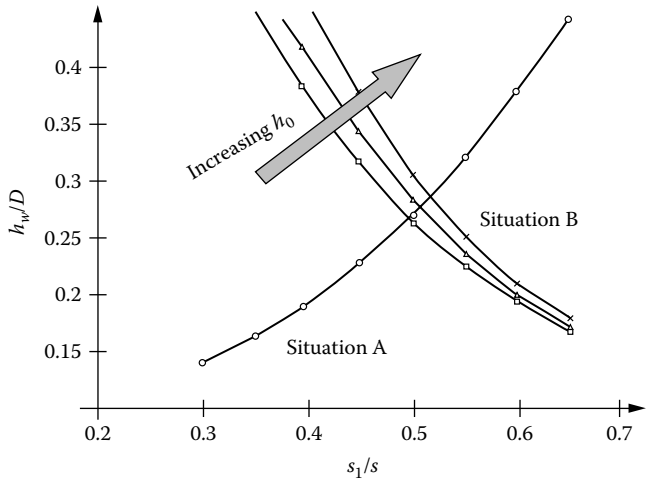


FIGURE 12.36 h_w/D in relation to s_1/s in comparison between front wheels (situation A) and rear wheels (situation B) encountering a step with a constant μ_a in situation A. (Reproduced from Hohl, G.H. and Peschak, B., A model to predict the overcoming of geometrically defined obstacles by a two-axle wheeled vehicle, *13th International ISTVS Conference*, TU München, Munich, Germany, 1999.)

12.3.2.3.3 Single and Rigid Wheel Model

Another way to approach the problem of overcoming a step is to consider only the front wheels and to summarize the traction forces F_R acting in the hub of the front wheel and coming from the rear wheels as in the case of a multi-axle vehicle [15,17].

Figure 12.37 shows the forces acting on a rectangular step at the very beginning of the overcoming. The wheel is driven by torque T_F. Force F_R is the sum of the traction forces of the rear wheels. Front axle load is W_F.

The equilibrium of forces and the torque in point A, respectively, in the center of the wheel are

$$T_F + F_R\left(\frac{D}{2} - h_w\right) - W_F l = 0 \tag{12.31}$$

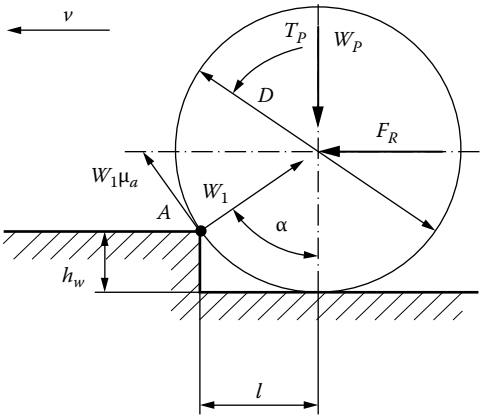


FIGURE 12.37 Single and rigid wheel model. (Reproduced from Antonov, A.S., Militärfahrzeuge Band 2, Militärverlag der Deutschen Demokratischen Republik, Berlin, Germany.)

$$T_F = W_1 \mu_a \times \frac{D}{2} \tag{12.32}$$

The sum of the forces in direction W_1 are

$$W_1 = F_R \sin\alpha + W_F \cos\alpha \tag{12.33}$$

The geometrical relationships are

$$l = \sqrt{Dh_w - h_w^2} \tag{12.34}$$

$$\sin\alpha = \frac{2l}{D} \tag{12.35}$$

$$\cos\alpha = 1 - \frac{2h_w}{D} \tag{12.36}$$

Equations 12.31 through 12.36 lead to the implicit equation:

$$W_F \sqrt{Dh_w - h_w^2} - \mu_a \left(F_R \sqrt{Dh_w - h_w^2} + W_F \frac{D - 2h_w}{2} \right) - F_R \left(\frac{D}{2} - h_w \right) = 0 \tag{12.37}$$

This rather complicated-looking equation can be solved and results in a relationship between h_w/D and the front wheel load W_F, the coefficient of traction μ_a in point A, the tractive forces coming from the rear wheels F_R, and the diameter of the front wheel D.

Equation 12.38 allows varying the different possibilities of the rear axle (e.g., multi-axle systems, different wheel loads, and different coefficient of traction on the front and rear wheels).

$$\frac{h_w}{D} = \frac{1}{2} \times \left(1 - \frac{1 - \mu_a \left(F_R / W_F \right)}{\sqrt{\left(1 + \mu_a^2 \right) \left(1 + \left(F_R^2 / W_F^2 \right) \right)}} \right) \tag{12.38}$$

The rules for a rigid wheel, described earlier, are the basis for the investigation of a flexible tire crossing a rectangular step. Taking the deformation of the tire into consideration, the forces do not act on the circumference of the ideal rigid wheel; they act as seen in Figure 12.38.

In this model, only the movements and the action and reactions of forces and momentums of the front wheel are considered. The tractive forces of the rear wheels are summarized in force F_R.

The process of crossing a step ($h_w < D/2$) by a wheel with a pneumatic tire can be divided into two stages. The movements from stage 1 to stage 2 are to be considered step by step:

$$N \sin\alpha + N\mu_a^* \cos\alpha - F_R - W_x = 0 \tag{12.39}$$

$$N \cos\alpha + N\mu_a^* \sin\alpha - W_F + W_Z = 0 \tag{12.40}$$

$$N\mu_a^* \left(\frac{D}{2} - \delta^* \right) + W_z \mu_a \left(\frac{D}{2} - \delta \right) - T_F = 0 \tag{12.41}$$

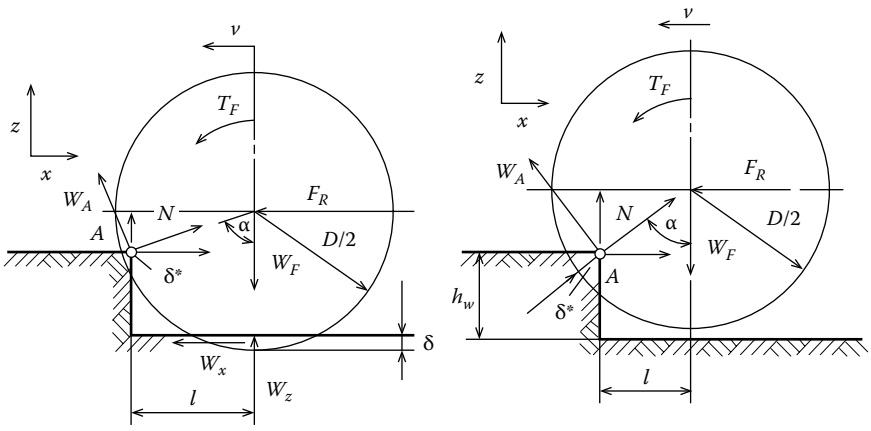


FIGURE 12.38 Single and flexible tire model. (Reproduced and modified from Ageikin, Ia.S., *Off-the Road Mobility of Automobiles*, Oxonion Press PVT. Ltd., New Delhi, India.)

where

μ_a is the coefficient of traction and the rolling plane

μ_a^* is the coefficient of traction between the tire and the edge of the step at point A

$\mu_a^* = 1,2\mu_a \ldots . - \ldots 1,3\mu_a$

$$h_w = \frac{D}{2}\left\{1 - \left[\frac{(1-(2\delta^*/D))}{\sqrt{1+((n-1)\mu_a - \mu_a^*/1-(n-1)\mu_a\mu_a^*)^2}}\right]\right\} \tag{12.42}$$

In the case of a front-wheel drive, the coefficient of traction force for the rear wheels μ_a has to be substituted by the coefficient of rolling resistance of the rear wheels μ_{aR}.

In the case of a rear-wheel drive, the coefficient of traction of the front wheels μ_a^* has to be substituted by the coefficient rolling resistance of the front wheels μ_{aF}^*.

12.4 SHOCKS AND VIBRATIONS

12.4.1 General

The surfaces over which off-road vehicles move are generally uneven because of the roughness and also because of single obstacles of the terrain. The crew and the vehicle experience more shocks and vibrations than ordinary road vehicles. Restrictions on speed can be imposed by small *irregularities* and a *wavy surface*, on the one hand, and larger obstacles, such as *bumps* and *ditches*, on the other.

Thus off-road speed is usually limited by the ability of the driver to withstand these vibrations in order to negotiate and retain adequate control of the vehicle motion (Figure 12.39).

The impact of vibrations to the crew depends on the following factors:

- Irregularities of the terrain
- Speed
- Suspension system of the vehicle
- Design of the seat

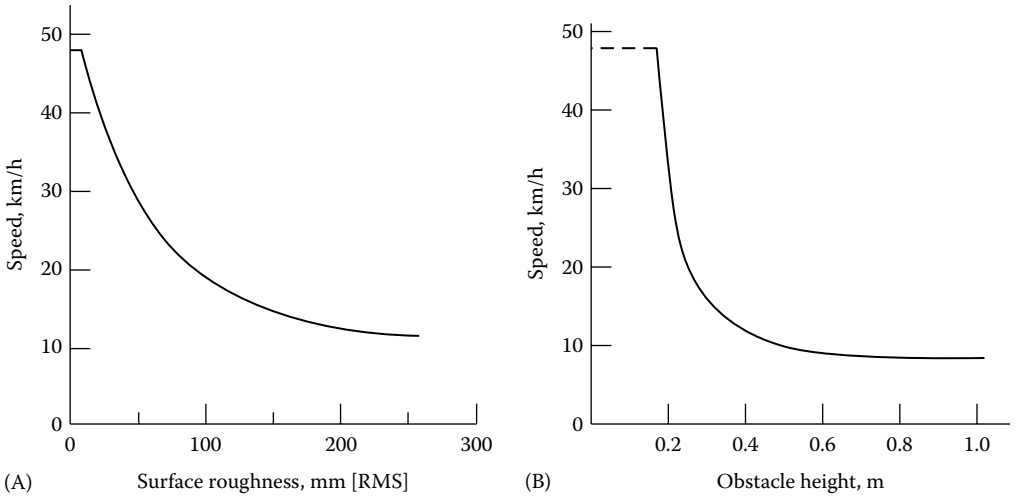


FIGURE 12.39 Speed of an MBT (A) versus surface roughness (B) versus obstacle height. (Reproduced and modified from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

Such developments to improve the driver seat, however, should not be at the expense of the driver's ability to "feel" the terrain. Therefore, a new definition of the term "speed" should be made:

S_{techn}: the technical speed is defined as the maximum speed that is possible from the technical performance of the vehicle and the available traction excluding the human factors.
S_{tol}: the tolerable speed is the speed that is tolerable for the crew. This speed will not be the same for all crew members.
S_{miss}: the mission speed or tactical speed of terrain vehicles should not be higher than the lowest tolerable speed for any crew member.
S_{fire}: the fire speed is the speed where a combat vehicle can successfully use the fire weapons and other equipment of the vehicle (Figure 12.40).

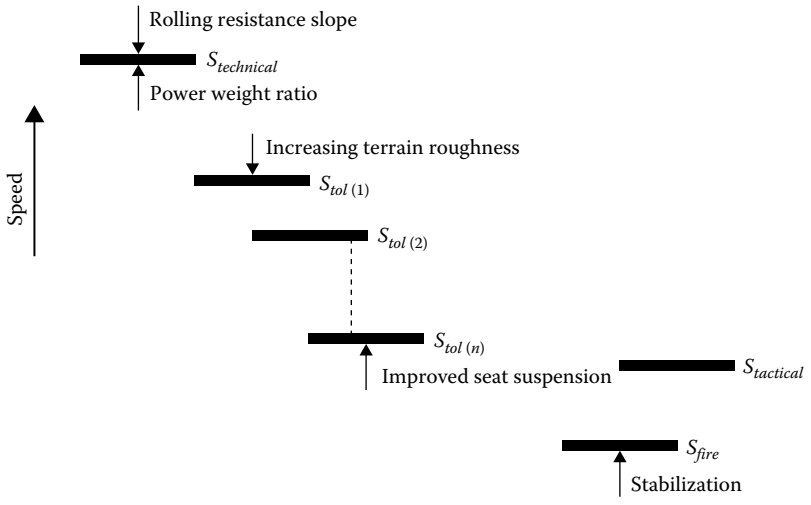


FIGURE 12.40 Speed definitions. (Reproduced from Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)

12.4.2 Human Response to Vibrations

12.4.2.1 General

The human body itself has no special sense organ to register mechanical vibrations nor does it have any specific defense mechanism to counter those vibrations. Prolonged exposure to whole body vibrations (WBV) can cause harmful physiological responses.

Both methods of investigation result in the conclusion that a human being exposed to mechanical vibrations can be affected by *reduction in comfort*, in *working efficiency*, and, in some circumstances, *health and safety.*

The reasons for human reactions to mechanical vibrations experienced by cross-country travel can be categorized into three groups. These are

- Physical effects
- Physiological effects
- Psychological effects

12.4.2.2 Mechanical Models of the Human Body (Physical Effects)

For the purpose of investigating the effects of mechanical vibrations, the human body should not be regarded as a rigid mass. Theoretical models of varying degree of complexity have been developed by several researchers to represent the human body as a mechanical system of masses, springs, and dampers (Figure 12.41).

Vibrations with frequencies below 1 Hz can also result in a general feeling of discomfort. These vibrations can cause kinetosis, or motion sickness, and the characteristics of these symptoms are quite different from those described earlier. Human reactions to vibrations below 1 Hz depend on a large number of external factors, for example, age, sex, vision, activity, and unpleasant odors (Table 12.2).

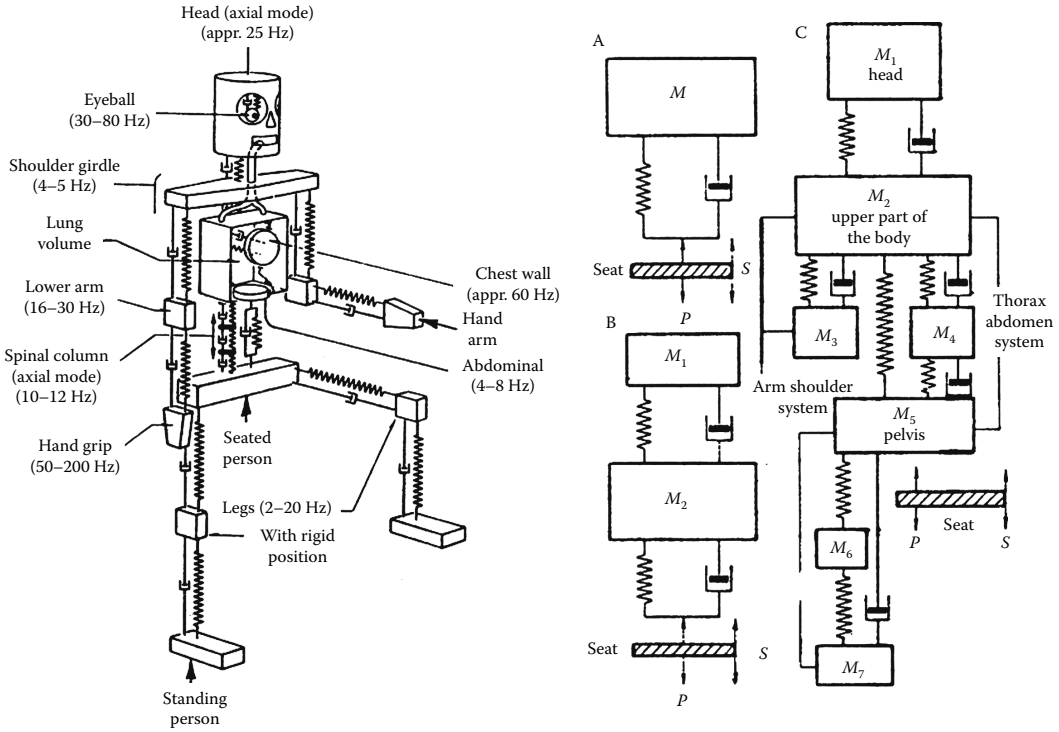


FIGURE 12.41 Mechanical substitute systems for the human body. (Reproduced from Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)

TABLE 12.2
Physiological Symptoms Dominate in the Range of 1–20 Hz

Symptoms	Frequency
General feeling of discomfort	4–7
Difficulties in breathing	2–4
Headache symptoms	13–20
Abdominal pains	4,5–10
Speech disorders	13–20
Urge to urinate	10–18
Lump in the throat	12–16
Chest pains	5–7
Increased muscle tension	4–9
Lower yaw symptoms	6–8

Source: Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.

12.4.3 Test Methods, Recommendations, and Standards

Many investigations have attempted to measure the subjective reactions to objective measurable vibrations and to develop methods for quantitatively describing short- and long-term effects of vehicle vibrations on occupants of terrain vehicles.

None of the methods is really satisfactory for off-road rides, as the criteria were developed for "boulevard" rides. The crew of off-road vehicles expects a harsh ride. Though the results vary in some cases, they still coincide with the fact that the subjective perception is a function of frequency. One problem is that the human reactions to WBV differ from person to person and that human beings cannot remember the intensity of vibrations very well.

Over the years, numerous ride comfort criteria have been proposed.

Two of these criteria are mentioned:

- "rms" acceleration criteria
- "Absorbed power" criterion

The test persons were subjected to a pair of sinusoidal vibration stimuli of the same frequency. All following stimuli are more intensive than the previous ones. The test person can select the acceleration in order to have the same subjective perception. All tests came to the result that the highest perception of discomfort was in the range of 4–8 Hz. This fact was the basis for the ISO 2631-1978 [31] and VDI 2057-1957 [32] standards.

In both standards, only the frequencies between 1 and 80 Hz are taken into consideration. Higher frequencies come near to the audio range and do not affect the discomfort described earlier.

Both standards put the equal sensation curves in three straight lines in a double logarithmic coordinate system and the same three criteria for the limits of human exposure to WBV (Figure 12.42):

- Preservation of comfort "reduced comfort boundary"
- Preservation of working efficiency "fatigue-decreased proficiency boundary"
- Preservation of health or safety "exposure limit"

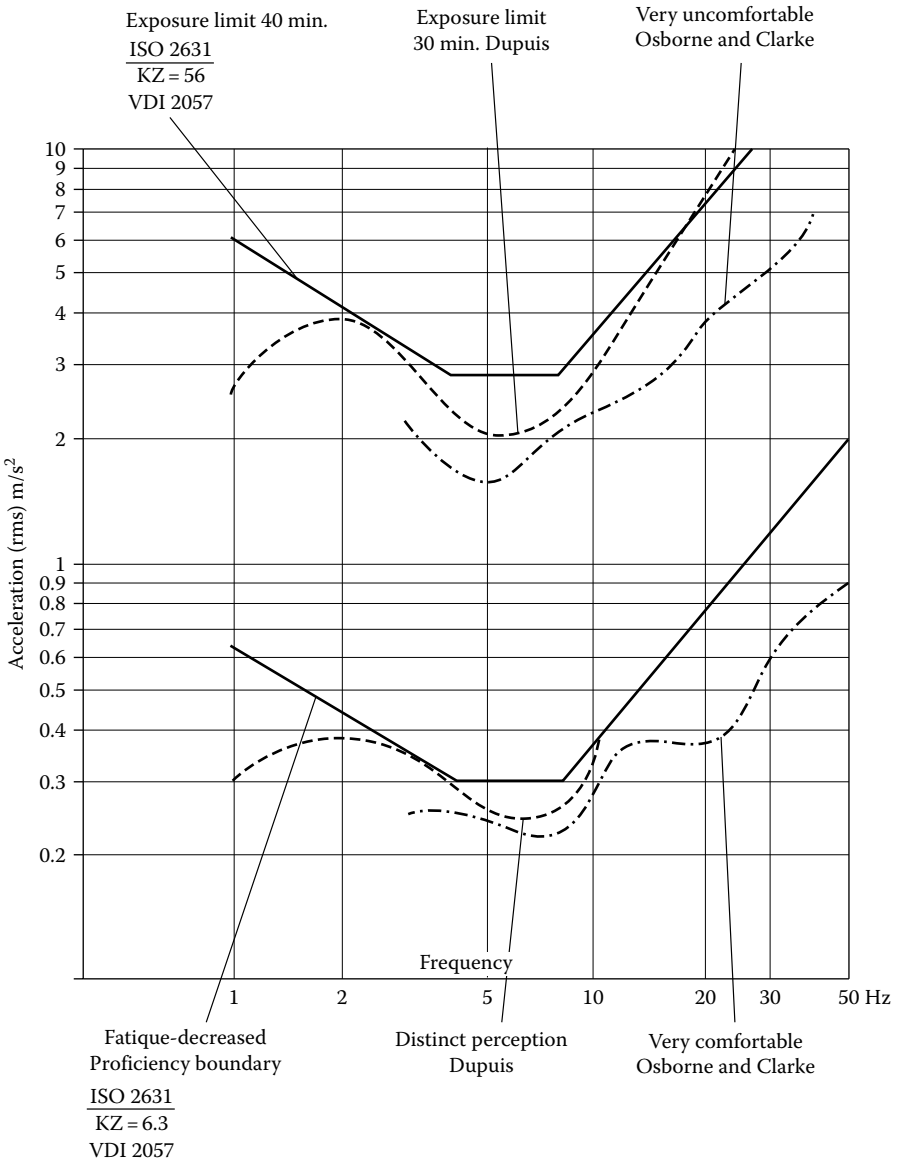


FIGURE 12.42 Equal sensation curves; comparison between standards and different research results. (Reproduced from Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)

Figure 12.43 shows clearly that the main influences on ride comfort are the speed and the surface. In all cases, the person on the rear seat of the loading platform was in the worst position.

The concept in the VDI 2057-1957 [31] and ISO 2631-1978 [32] standards was not supported by research results in laboratory tests. The modified and new standards ISO 2631-1:1997 (E) [33] and ÖNORM ISO 2631-1: 2005-11-01 [34] extend the frequency range below 1 Hz and delete the concept of "fatigue-decreased proficiency boundary" due to vibration exposure.

In the earlier-mentioned standards, methods for the measurement of periodical, random, and transient WBV are defined.

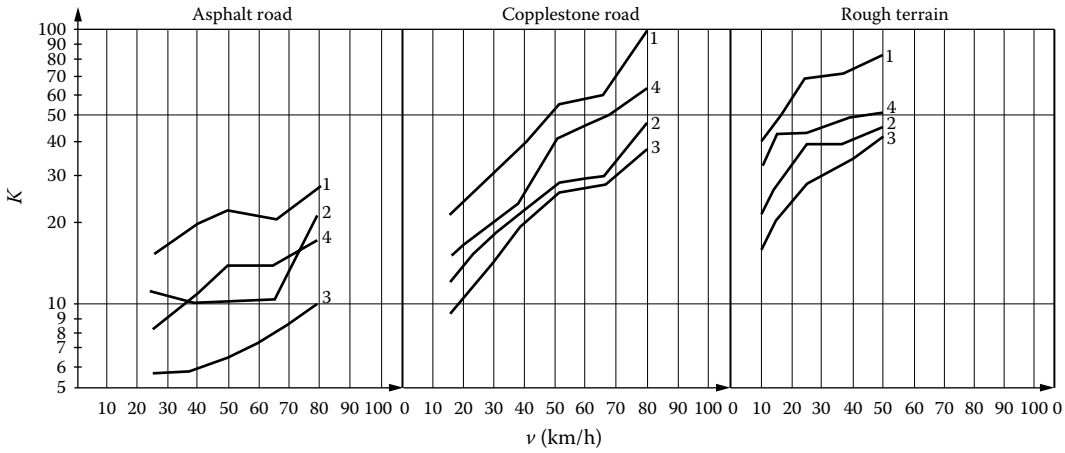


FIGURE 12.43 *K*-values depending on surface, seat position, and speed. (Reproduced from Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.) Position 1: rear seat; position 2: front seat of the loading platform/commander seat; position 3: seat mount of the driver's seat; position 4: driver seat.

The frequency range considered is

- 0.1–0.5 Hz for motion sickness
- 0.5–80 Hz for health, comfort, and perception (Figure 12.44)

W_k is the frequency weightings for a sitting person in z-direction; W_d in x, y direction for a laying person; and W_f for kinetosis.

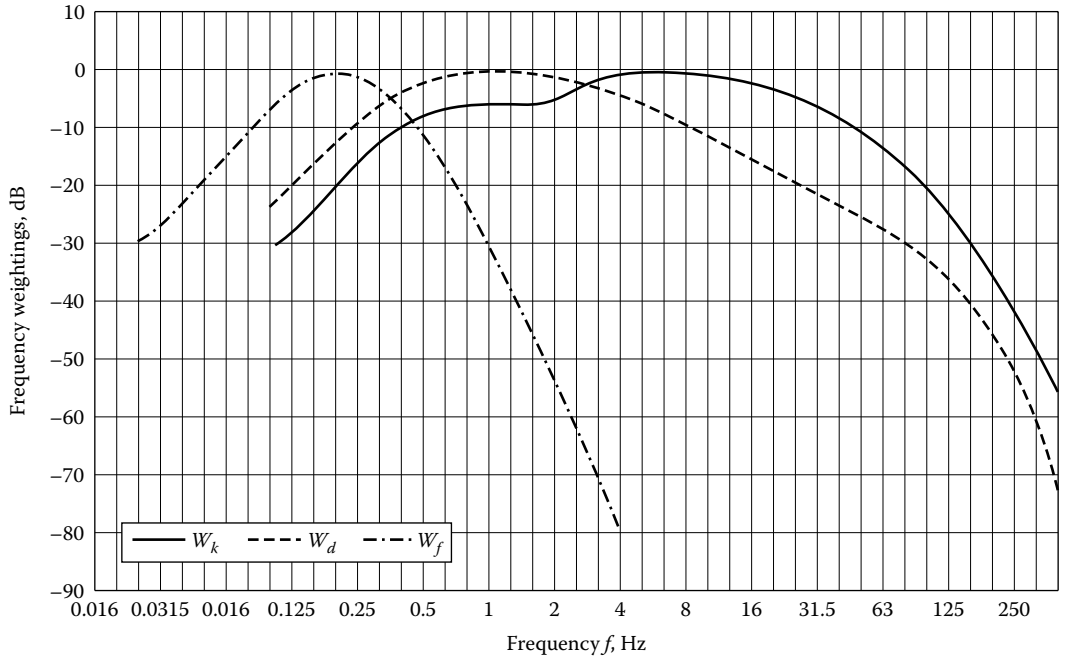


FIGURE 12.44 *K*-values depending on surface, seat position, and speed. (Reproduced and modified from Hohl, G., Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)

12.4.3.1 Absorbed Power

The U.S. Army uses "absorbed power" to measure the level of discomfort of an occupant riding in a vehicle. Average absorbed power can be calculated if the time history of acceleration at a selected point, for example, the driver's seat, is known. The smaller the absorbed power is for a given vehicle, running over a specific profile at a certain speed, the better the vehicle ride [35,36].

The absorbed power concept offers certain advantages over pure acceleration for terrain vehicle ride quality evaluation. Firstly, it takes into account multifrequency and multiaxial vibration across a broad frequency range. Secondly, it provides proper weighting functions for all frequencies and axes according to body response.

12.5 TECHNICAL CONCEPTS

This section describes technical solutions and concepts of wheeled and tracked off-road vehicles in armored and non-armored versions.

12.5.1 Vehicle Bodies

The vehicle bodies for off-road vehicles in tracked and wheeled versions are

- Frame
- Hull

12.5.1.1 Frame

Frames used for non-armored vehicles are

- Ladder frame (Figure 12.45)
- Central tube frame (Figure 12.46)

Vehicles transporting closed van bodies and containers across terrain conditions with high obstacles have to incorporate a frame-to-body connection that could never be reached with a non-torsion-resistant frame [37,38] (Figures 12.47 and 12.48).

The cross section on the right shows an overall view of a torsionally resilient frame of an 8 × 8 terrain truck. The frame consists of two side members and several tubular cross-members that are mounted in holes in the side members and welded to their inner and outer side.

Calculations and tests have demonstrated that when lifting one wheel of a torsion-friendly frame of 1 m, the frame handles no more than 57% of total torsion contrary to the stiff frame, which

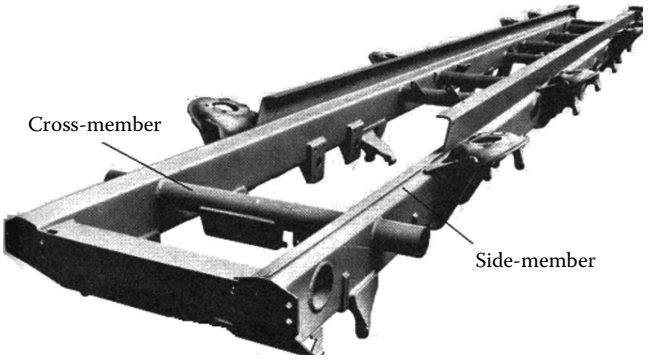


FIGURE 12.45 Conventional ladder frame. (Reproduced and modified from Klanner, R., Cross-country trucks for fast off-highway operation, *6th International ISTVS Conference*, Vienna, Austria, 1978.)

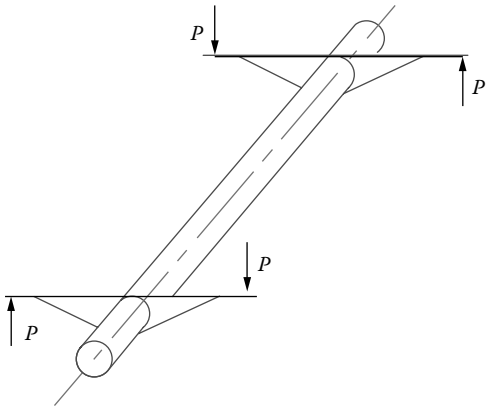


FIGURE 12.46 Central tube frame. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

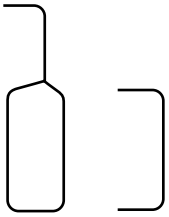


FIGURE 12.47 Closed and open side member sections of a ladder frame (left). (Reproduced from Klanner, R., Cross-country trucks for fast off-highway operation, *6th International ISTVS Conference*, Vienna, Austria, 1978.)

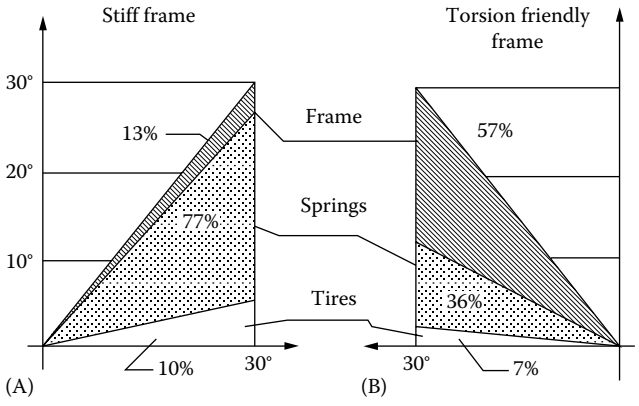


FIGURE 12.48 Deflections of a stiff (A) and a torsion-friendly frame (B). (Reproduced and modified from Hoepke, E. et al., *Nutzfahrzeugtechnik ATZMTZ-Fachbuch*, Vieweg Verlag, Berlin, Germany.)

covers 13% of the deflection. In a stiff frame concept, the springs do the main work in twisting situations. Central tube frames and *X*-frames are extremely stiff, but only the first concept is realized in practice.

12.5.1.2 Hull

The arrangement of the driveline and the design of the hull are dependent on the function of the armored vehicles. The main representatives of this type are MBTs, APC, and AWV.

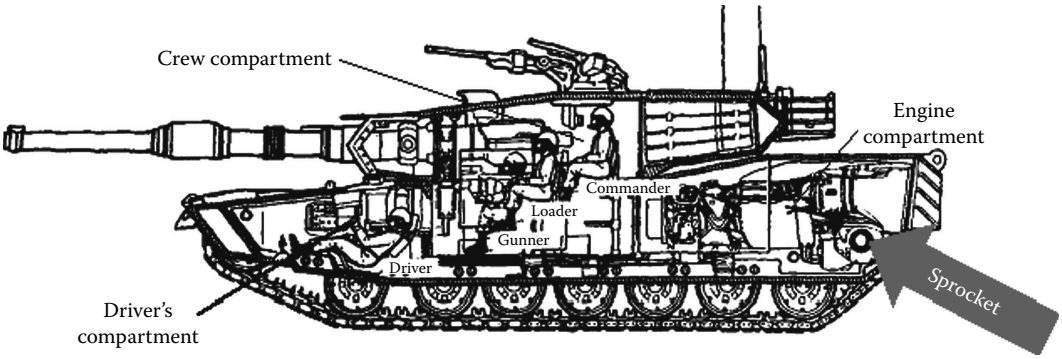


FIGURE 12.49 Space partition of an MBT (sprocket in the rear). (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

In an MBT, the engine compartment and the final drive are located in the rear. The crew compartment for the commander, the gunner, and, in some cases, the loader is in the middle and the driver is in the front part.

The task of an APC is mostly to cooperate with MBTs and to transport an armor-protected group of six to nine infantrymen. The infantrymen are located in the rear of the hull. The commander and the gunner are in the middle, and the driver and the engine–transmission block are in the front and thus the final drive is also located there.

Most of the other armored tracked and wheeled vehicles with other tasks belong to one of the two principal concepts described earlier (Figure 12.49).

The Israeli MBT "Merkava" is an exception; it can accommodate a small group of infantrymen in the rear of the tank, which means that the engine is in the front of the vehicle.

This configuration offers the advantage that there is no need for an additional APC to perform the infantry functions. Moreover, the front of the vehicle is well protected by the engine block. This MBT is a front-droop vehicle and its sight instruments are negatively influenced by the flickering hot air from the cooling system of the engine (Figure 12.50).

The most frequently used material for the hull, in order to increase its resistance to penetration by projectiles, is heat-treated and hardened steel. Aluminum armor began to develop in the United States in 1959 with the production of an aluminum-hulled personnel carrier, which became the M113, the most used APC in the world, but the latter material remained a niche product.

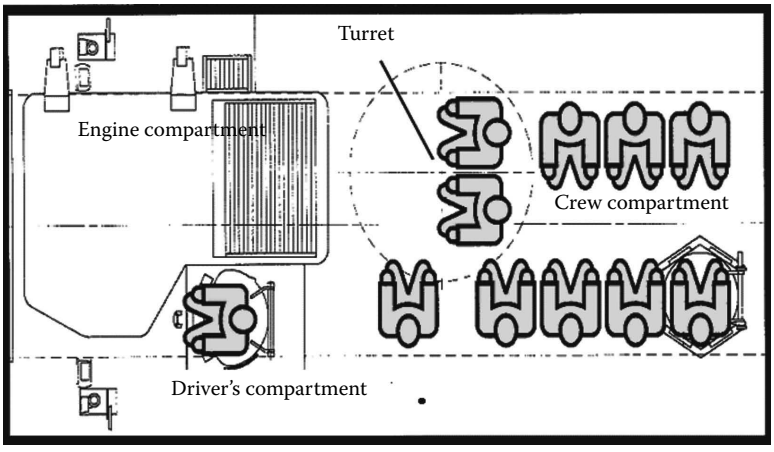


FIGURE 12.50 Space partition of an APC (sprocket in the front). (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

The advantage of a hull compared to a frame is its more rigid structure as the moment of resistance increases by the square of the distance to the neutral axis.

12.5.2 Power Train

12.5.2.1 General

The arrangement and the components of the power train differ for wheeled and for tracked off-road vehicles (Figures 12.51 and 12.52).

Wheeled Vehicles	Tracked Vehicles
Engine	Engine
Gearbox	Transmission–steering unit
Transfer case	Final drive unit
Axle differentials	Tracks
Final drive unit	
Wheels	

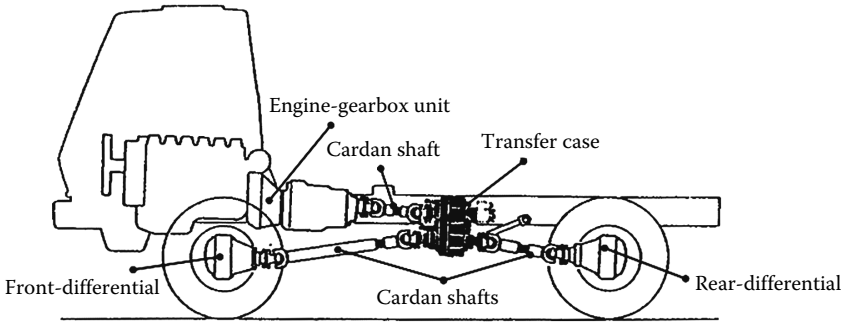


FIGURE 12.51 Conventional power train of a wheeled non-armored vehicle. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

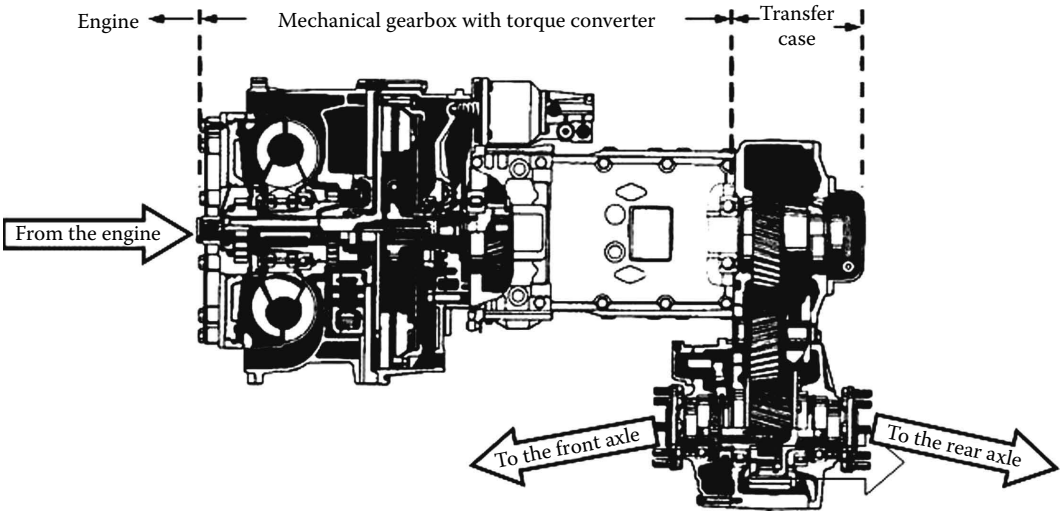


FIGURE 12.52 Engine–gearbox transfer case unit concept. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

There are also exceptions regarding these two concepts mentioned earlier. The light military vehicle "Iltis" has no transfer case. The front-wheel drive comes directly from the gearbox. For off-road operations, there is a very low first gear speed installed. This concept requires only one cardan shaft between the engine–gearbox unit and the rear-wheel drive.

Due to the dependent connection between the engine–gearbox block, the long distance between the front axle and the front engine hood contour, the vehicle has a poor angle of approach [39].

12.5.2.2 Engine

It should be evident that acceleration increases with power-to-weight ratio, in spite of the fact that the maximum possible tractive effort is limited in all cases to the same value by the coefficient of traction. Thus, the higher the power-to-weight ratio, the lower is the required overall gear ratio, which has the threefold effect of reducing the number of gears that are necessary, of increasing the average value of the tractive effort over the speed range, and of reducing the value of the inertia mass factor, which clearly increases the acceleration.

12.5.2.2.1 Diesel Engine

The reason for the general adoption of diesels in preference to spark-ignition engines is that they increase the operating range for a given fuel tank capacity because of their greater thermal efficiency and, therefore, lower fuel consumption. For diesel engines, the injection system and enhanced exhaust gas aftertreatment are the subject of intensive research.

The actual acceleration of two typical tanks with different power-to-weight ratios is shown in Figure 12.53 in terms of curves of velocity and distance covered against time. The tank with the higher power-to-weight ratio can obviously attain a given velocity or reach a particular distance more quickly than the other.

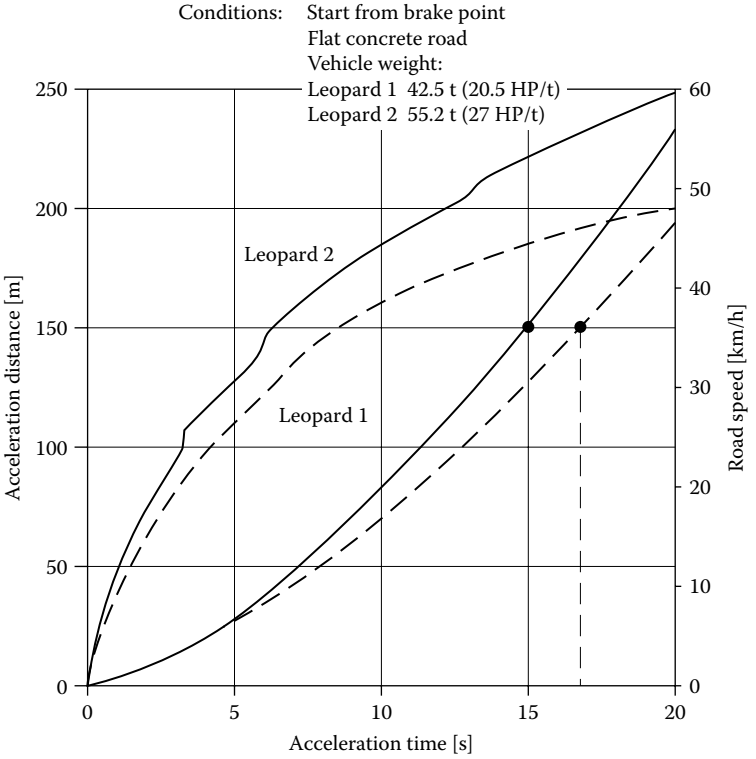


FIGURE 12.53 Tank acceleration "Leopard 1" versus "Leopard 2." (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

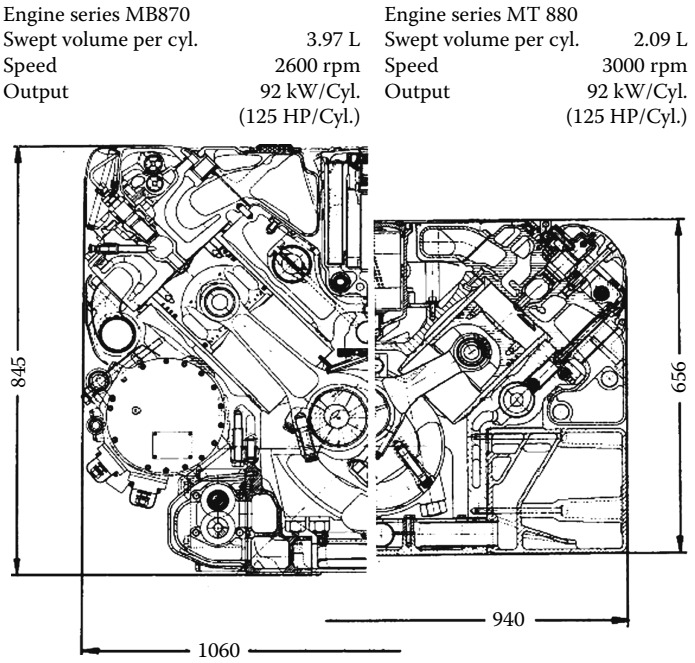


FIGURE 12.54 Comparison of the size of the MB 870 engine and MB 880 engine. (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

The situation for tracked armored vehicles is quite different, as these vehicles have limited space in their engine compartments [40–42]. Such engines are mostly based on commercial models but have to be adapted for a small volume. For these tracked armored vehicles, the engines with accessories, air-cooling and filter systems, as well as a transmission–steering system are integrated into a power pack. This power pack must be tailored to the specific requirements. Producers of engines for these types of vehicles are trying to reduce the dimensions. A good example for this trend of the downsized engine is the MTU 870 modified to the MTU 880 with the same power (Figure 12.54).

In recent years, the power-to-volume ratio kW/m^3 has improved rapidly and will presumably increase further.

Only for heavy tracked vehicles, the question arises: Diesel Engine or Gas Turbine?

Soon after "diesels" had become generally accepted as tank engines, an alternative to them began to emerge in the form of gas turbines. The advantages of gas turbines are the high specific output and the favorable torque–speed characteristics. Another advantage of the gas turbine over a conventional combustion engine with translation motion of the pistons is that the rotating motion of the turbine is transferred into another rotating motion.

The main reason that gas turbines are not more frequently used in terrain and also in other vehicles is the fact that the fuel consumption of a gas turbine in a part-load situation is very disappointing. Even under full load, the fuel consumption exceeds that of a diesel engine. The thermal process of a gas turbine needs an enormous surplus of air. This needs a bigger filter system and grill inlets, which is problematic for armored vehicles.

The low emissions of carbon monoxide and unburned hydrocarbons make the exhaust gas quality of turbines better than for diesel engines. Service intervals and maintenance requirements are positive features.

The AGT-1500 used in the U.S. MBT "Abrams" (Figure 12.55) is one of the few examples of gas turbines in vehicles (Figure 12.56).

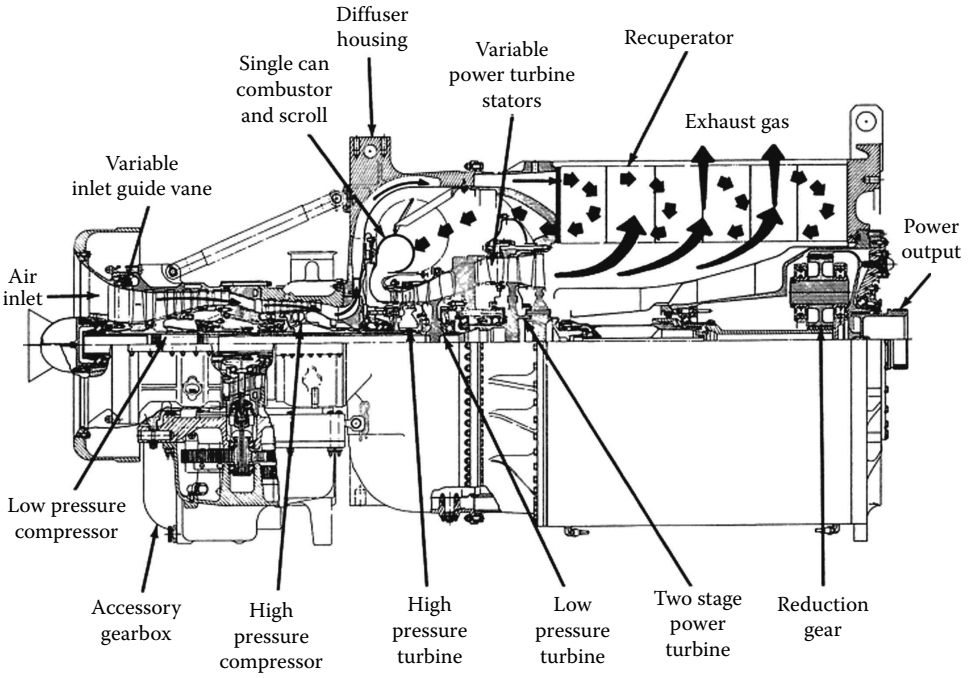


FIGURE 12.55 AGT-1500 for the U.S. MBT "Abrams." (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

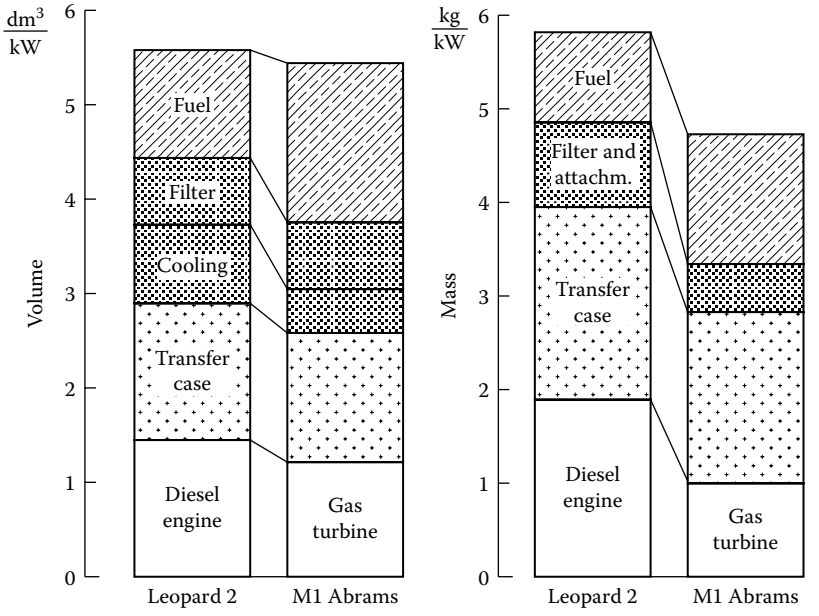


FIGURE 12.56 Differences of masses and volumes between diesel engine and gas turbine. (Reproduced and modified from Schmid, I., Grundlagen der Geländegängigkeit—Wechselwirkung zwischen Rad und nachgiebigen Boden, Vorlesung des Institutes für Kraftfahrwesen und Kolbenmaschinen der Universität der Bundeswehr Hamburg, Hamburg, Germany, 1990.)

12.5.2.3 Transmission

12.5.2.3.1 Wheeled Vehicles

For wheeled off-road vehicles of all concepts, the standard question arises whether to use a manual or an automatic transmission. Both systems are in use. However, in order to release the driver from routine functions, the number of automatic transmissions for off-road vehicles increases.

The idea of an increased number of gear steps leads to a *C*ontinuously *V*ariable *T*ransmission (CVT) [43,44]. This system, based on a metallic band used for passenger cars, is not suitable for the high torques needed for terrain vehicles. CVTs for this purpose are based on the splitting of power from the drive unit into a mechanical and a hydraulic part. Until now, this system has only been used in tractors (Figure 12.57).

The benefits of a CVT system are

- Fully automated power train from the engine to the wheel
- Smooth vehicle acceleration
- Very good efficiency of the engine as the engine speed can be adjusted independent of the vehicle speed
- No perceptible gear steps and no interruption of the power flow

12.5.2.3.2 Tracked Vehicles

Steering principles of tracked vehicles have certain unique features and are quite different from those of wheeled vehicles. Separate treatment of the steering of tracked vehicles is, therefore, required. Nowadays, there are two possible methods that can accomplish the steering of tracked vehicles. These are *skid steering* and *articulated steering*.

Skid steering is the most used method. Speed of one track is increased and that of the other is reduced, in order to create a turning movement (Figure 12.58).

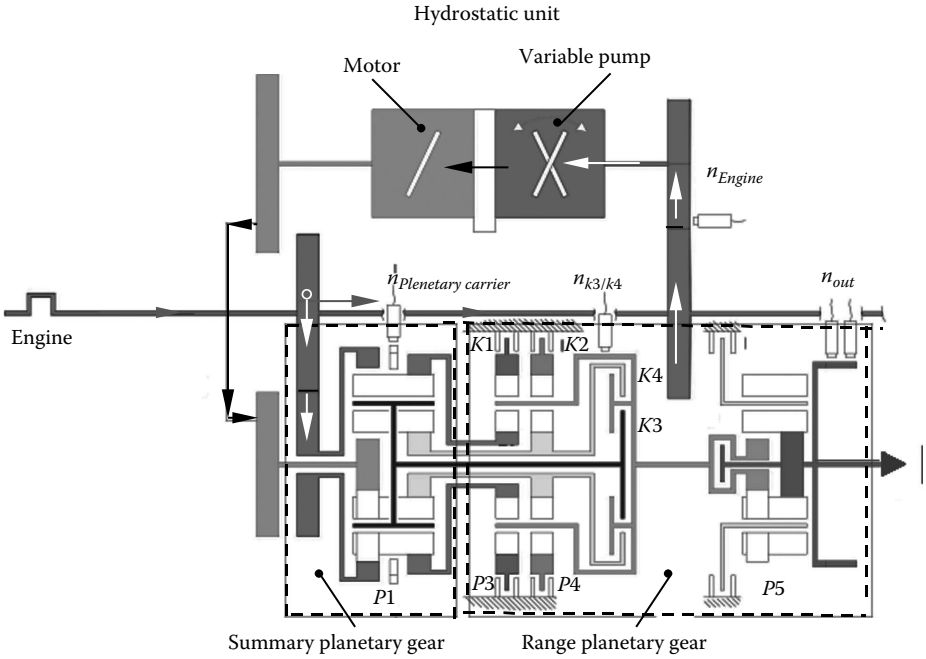


FIGURE 12.57 Mechanical–hydrostatic CVT; principle. (Reproduced from Aitzetmüller, H., Steyr S-matic—The future continuously variable transmission for all terrain vehicles, *13th International ISTVS Conference*, Munich, Germany, 1999.)

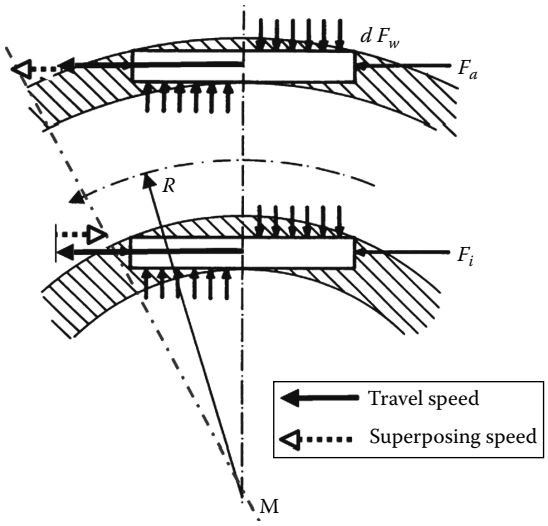


FIGURE 12.58 Principle of skid steering. (Reproduced and modified from Merhof, W. and Hackbart, E.M., *Fahrmechanik der Kettenfahrzeuge,* Leuchtturm, Germany.)

Modern steering units for tracked vehicles add an additional velocity to one track and subtract the same amount of velocity from the other. These additional velocities are done either mechanically (U.S. MBTs M47, M48, M60 [45] or by a hydraulic pump–motor unit, Figure 12.59). The final drive consists of a planetary gear assembly. The additional velocities are usually inserted via the sun wheel. An additional gear wheel on one side causes movement in the opposite directions of the left and the right track.

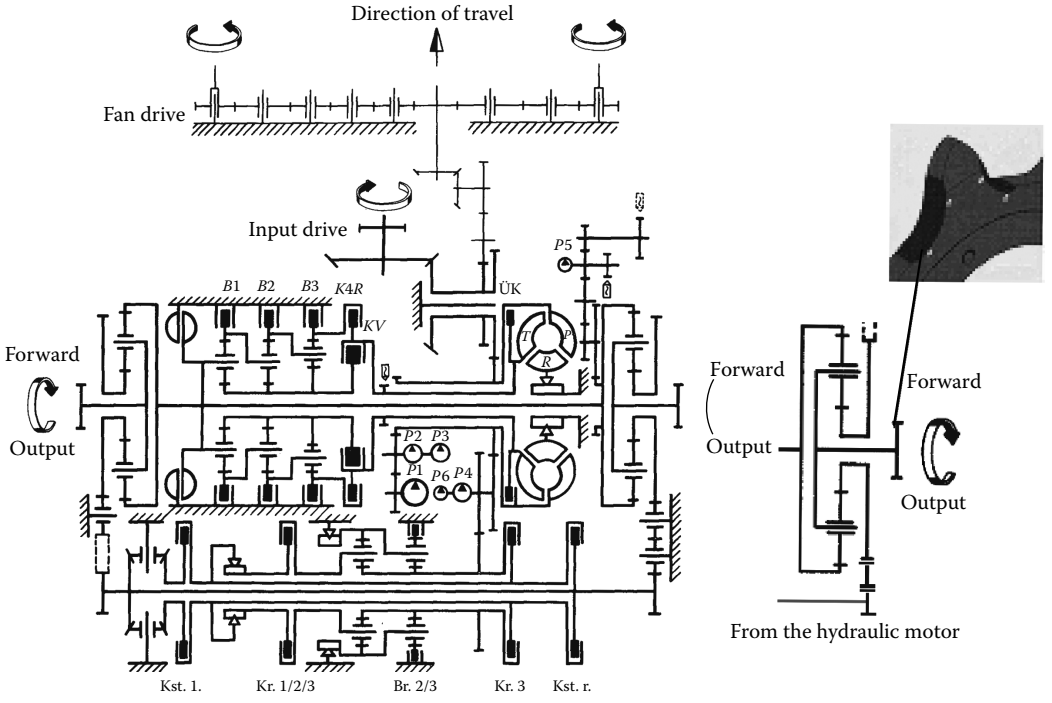


FIGURE 12.59 Transmission–steering unit. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria; NN, Brochure of ZF Company, Friedrichshafen, Germany.)

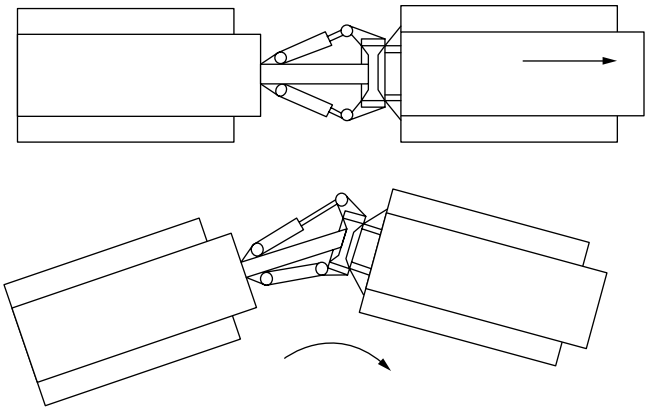


FIGURE 12.60 Principle of an articulated steering. (Reproduced from Wong, J.Y.: *Theory of Ground Vehicles*. 2001. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

In articulated steering, turning is initiated by activating the steering joint between the two units, and no adjustment of the thrusts of the outside and inside tracks is required. Thus, the resultant forward thrust of the vehicle can be maintained during a turn. Articulated steering can, therefore, provide tracked vehicles with better mobility during turning maneuvers than skid steering, particularly over soft ground. But on the other hand, a pivot turning is not possible. The most prominent soft-skinned vehicle is the Swedish "BV 206" (see Figure 12.60).

The kinetic of skid steering is described in detail in Ref. [7].

12.5.2.4 All-Wheel Drive and Differential Locks

Both all-wheel drive and differential locks are essential elements of wheeled off-road vehicles [46]. All-wheel drive technology has established itself for the following reasons:

- To increase climbing performance
- To increase traction through full utilization of static friction
- To improve moving of traction
- To increase payload and trailer load

Two principal all-wheel drive systems are in use:

- Selectable
- Permanent

Selectable all-wheel drives can be actuated mechanically by a disc clutch for light vehicles or a dog clutch for heavy vehicles [46] with high torque. The actuation can be done either *manually* intentionally by the driver or *automatically* in the case of an electronic *A*utomatic *D*rivetrain *M*anagement (ADM) [47–50] or by a *viscous coupling.*

The transfer case has two functions:

- To distribute torque to all wheels
- To increase torque for terrain missions by a two-stage gearbox

To compensate the speed differences of the axle drives, permanent all-wheel drive systems have differentials between the axles. These can be bevel gears or planetary gear differentials.

In the case of bevel gears central differentials, the torque distribution is fixed at 50:50. With planetary gear differentials, the drive torque can be split to the two drive axles, as required, by selecting the ratio. Typical torque distribution ratios between front and rear axle are 33:66.

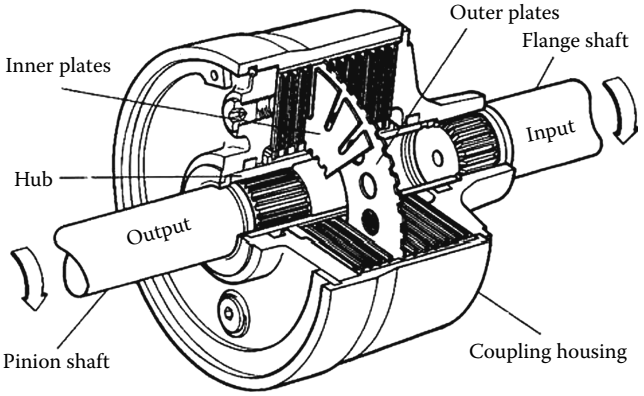


FIGURE 12.61 Viscous coupling; view. (Reproduced and modified from Braess, H.-H., *Handbuch Kraftfahrzeugtechnik*, Vieweg Verlag, Berlin, Germany.)

The use of a *viscous coupling* [51] provides another way of linking two axles, building up the coupling torque between the front and rear axles through viscous friction (slip controlled) depending on the speed differential between the front and rear axles. The viscous coupling is an encapsulated multi-plate system with high-viscosity silicon fluid and automatically distributes the torque between the front and rear axles as a function of the traction conditions at the wheels. If the speed difference is very small, only the rear wheels are driven. Should there be an increasing difference between RMS of the front and the rear wheel drive, the silicon increasingly stiffens with the result that the locking torque between the plates increases and an infinite drive is connected to the front wheels, extending as far as a rigid connection.

The level of coupling torque depends on the clutch characteristic selected. It can be influenced by the level, viscosity, and temperature of the oil (using silicon oil with low viscosity change with temperature) (Figures 12.61 and 12.62).

12.5.2.4.1 Differential Locks

The longitudinal (in the case of all-wheel drive) and transverse differentials have the function to adjust independently of each other according to the different speed of the wheels of the different axles and the difference of the speed by the left and right wheels according to the different distances travelled.

Differential mechanisms can be installed into transfer boxes (longitudinal splitting of speed and torque), as described earlier, and differentials (transverse splitting) between the wheels of one axle.

During many driving maneuvers, windup torques occur in power trains with rigid all-wheel drive, positive differential locks, or nonpositive differential locks with high lock up ratio. These self-generated torques are mainly caused by differing dynamic tire diameters, kinematic slip between front and rear

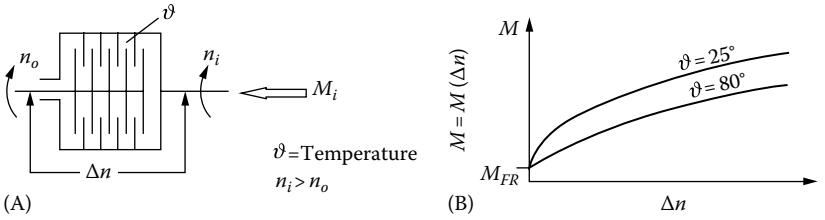


FIGURE 12.62 Viscous coupling; principle (A); torque M versus speed difference Δn (B). (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

axle during cornering, and cinematic slip between the wheels of one axle. The phenomena of self-generated torques are described in detail in Ref. [52].

In extreme off-road situations, there is a need for locking the longitudinal and transverse differentials. This can be done in different ways:

- By activating manually or automatically and by mechanical, magnetic, pneumatic, or hydraulic means and blocking any compensating action by locking the differential unit with a dog clutch. This makes the axle up to a certain degree rigid, with all the resulting advantages and disadvantages.

  The use of such a traction control system is thus appropriate where there is inadequate traction for one wheel or one axle and should only be temporary. This system needs a dog clutch.
- By using mechanical self-locking differentials, which are also known as limited-slip or locking differentials. These are differentials with a compensating action that is deliberately tight and restricted by means of their mechanical abilities. This enables them to transmit torque to one wheel or axle even when the other wheel or axle is spinning because of poor grip. Locking differentials are divided into load or torque controlled and speed or slip controlled.
- By a viscous coupling parallel to the differential. (The principle of a viscous coupling is described earlier.)

The mostly used mechanical differential locks are

- Multi-disc clutch self-differential lock
- TORSEN self-locking differential
- Dog clutch (Figure 12.63)

Multi-disc clutch self-differential lock: The locking effect of a self-locking differential with multi-disc clutch relies on the torque-dependent internal friction generated in two multi-disc clutches mounted symmetrically in the differential cage. The self-locking action results from a combination of the load dependency and spring loading of the multi-disc clutches.

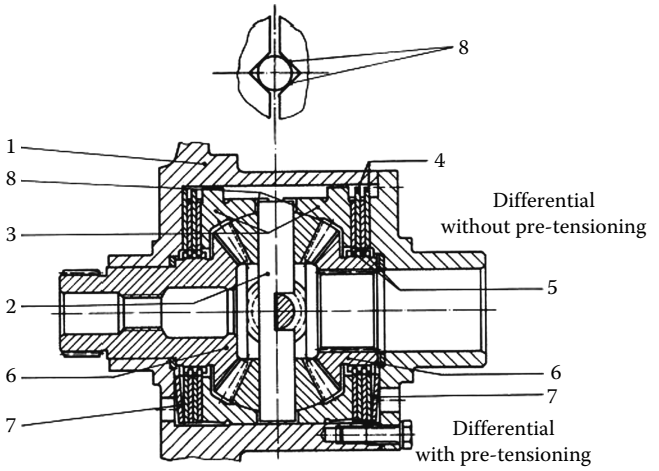


FIGURE 12.63 Limited-slip differential with prestressed multi-disc clutches. (Reproduced from Lechner, G. and Naunheimer, H., *Automotive Transmissions*, Springer Verlag, Berlin, Germany, *Anti-Wheel-Spin Systems for Road Vehicles*, London U.K., 1988.)

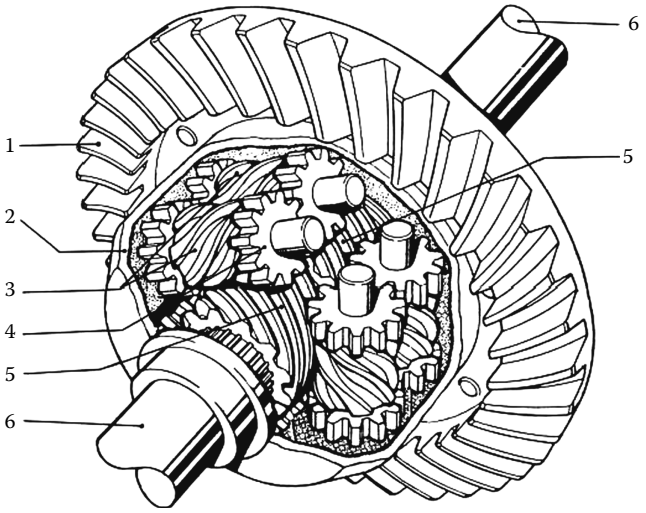


FIGURE 12.64 TORSEN self-locking differential with a worm wheel. (Reproduced from Schlechter, H.P., Design and performance of central tire inflation systems, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

TORSEN self-locking differential: This kind of differential with self-locking abilities can be used for longitudinal and transverse differentials. TORSEN stands for "*tor*que *sen*sing" and is based on the mechanical properties of a worm and bevel gear interaction. The principle is described in detail in Ref. [53] (Figure 12.64).

12.5.2.5 Final Drive

High torque needs high volume and thus weight for gear reductions and shafts. Therefore, the intention of a designer of an off-road vehicle is to go through the vehicle with as low torques as possible but at the end of the power train, high torque is needed to overcome all the resistances of a terrain travel. Therefore, many off-road vehicles have an additional gear reduction in the final drive unit. There are two solutions available:

- Spur gear hub drive
- Planetary set

For soft-skinned wheeled vehicles, spur gear drive reduction can also be used for portal axles (e.g., "Pinzgauer" and "Unimog"). Portal axles have the advantage of an increase in ground clearance but also the disadvantage of a higher C.G. The spur gear concept is also used for tracked vehicles as the Austrian light tank destroyer "Kürassier" shows.

In the case of a planetary unit, it can be located in the hub for wheeled vehicles and for tracked vehicles in the sprocket assembly (Figure 12.65).

12.5.2.6 Tires

In off-road operations, various types of wheeled vehicles are expected to operate over a great variety of surfaces in cross-country transport [54].

During the last years, more military tasks have been transferred from tracked armored vehicles to wheeled armored vehicles, as studies show that in many cases off-road mobility can be obtained by wheeled vehicles. However, one cannot deny that a certain amount of hesitation regarding the suitability of vehicles with pneumatic tires lies in the vulnerability of the tires, as they are very often exposed to the effects of enemy attack. Therefore, military wheeled off-road vehicles are often provided with such an *emergency running* or *run-flat* system.

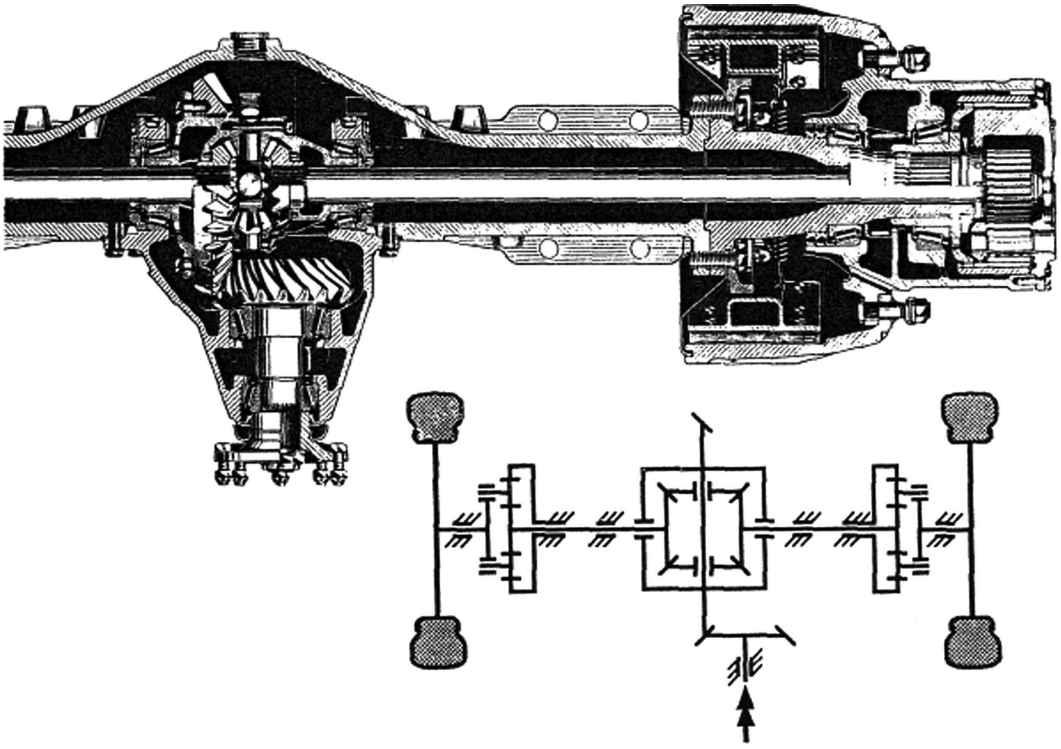


FIGURE 12.65 Final drive planetary gear for a wheeled vehicle. (Reproduced from Schlechter, H.P., Design and performance of central tire inflation systems, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

12.5.2.6.1 Tire Structure

The principal tire structure of off-road tires can be divided into two main groups [7,54]:

- Bias-ply (Diagonal) tires
- Radial-ply tires (Figure 12.66)

12.5.2.6.1.1 Bias-Ply (Diagonal) Tires These tires have two or more body plies that cross at a crown angle of 40° to the center line of the tread and strengthen both the sidewall and

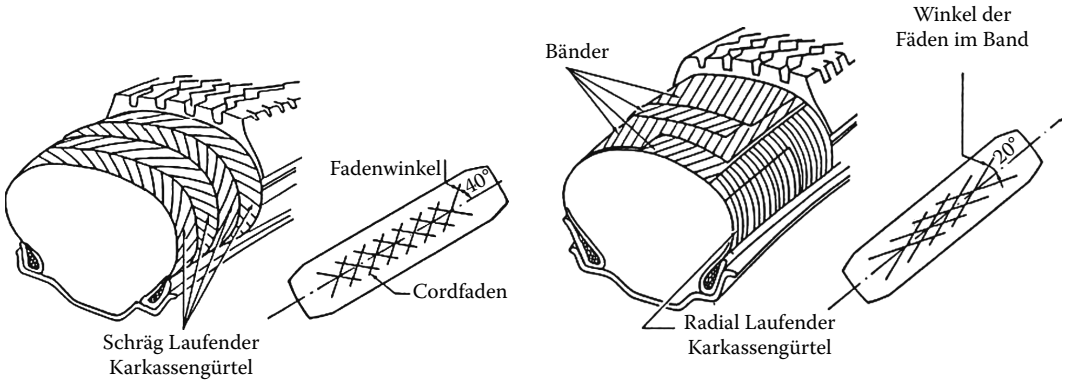


FIGURE 12.66 Tire constructions. (Reproduced from Wong, J.Y.: *Theory of Ground Vehicles*. 2001. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.)

the tread area. Bias-ply tires will always have an even number of body plies because of this. A bias-ply tire usually has a rounded tread face and even at low pressures will not have any sidewall bulge [7].

12.5.2.6.1.2 Radial-Ply Tires The construction of a radial differs from a bias-ply tire: The body cords run across the tire perpendicular to the beads. Radial tires have belt plies that run almost circumferentially around the tire, under the tread, to stabilize the tread, to strengthen the tread area, and to add flexibility to the sidewall. By restricting tread movement during contact with the ground, the belt plies improve tread life.

A radial tire usually has a squared-off tread face and a distinctive sidewall bulge at the tire contact point, even with proper inflation.

While some radial tires may have "tread" on the sidewall, this tread usually has a break at the sidewall transition and the sidewall tread is very thin to minimize heat buildup from radial sidewall flexing.

12.5.2.6.2 Special Tire Systems

*T*ire *I*nflation *S*ystems (TIS) [55–57] and tires with emergency capabilities (run-flat systems) [58–60] will be standard in the future.

It is well known that one possibility to enhance the mobility on soils with limited load carrying capacity is to reduce the tire pressure. This measure helps to increase the transferable tractive force. This is due to the enlargement of the tire tread area, an improved adaptation of the tire footprint to the soil, and enlargement of the effective tire diameter and thus a reduction of the rolling resistance coming from soil compaction (Table 12.3).

The great influence of tire pressure on the transferable tractive force can be seen in Figure 12.67. A tire pressure reduction from 3.5 to 2.0 bar results almost in a duplication of the DP.

The purpose of a run-flat tire system is to allow a vehicle to be driven with one or more flat or damaged tires over a certain distance. A run-flat system is not a replacement or a substitute for a tire.

A run-flat tire is a pneumatic vehicle tire that is designed to resist the effects of deflation and to enable the vehicle to continue to be driven, depending on the type of tire, at reduced speeds (i.e., 80 km/h or 50 mph) and for limited distances.

There are two basic run-flat technologies currently available, described in the following [56]:

12.5.2.6.2.1 Self-Supporting The tire is built with stiffer side walls that can bear the weight of the vehicle even when the pressure within the tire is greatly reduced. The side walls are typically constructed using layers of rubber and a heat-resistant cord that prevent the side walls from folding or creasing. The bead around the edge of the tire is also specialized to grip the wheel rim so as to avoid becoming detached from the rim.

TABLE 12.3
Tire Foot Prints of a Tire with 24.1 kN Load

Tire pressure (bar)	6.2	4.1	3.1	2.1
Tire print length (cm)	25.8	32.0	33.8	36.3
Deflection (%)	8.5	15	23	30

Source: Reproduced and modified from Schlechter, H.P., Design and performance of central tire inflation systems, *6th European ISTVS Conference*, Vienna, Austria, 1994.

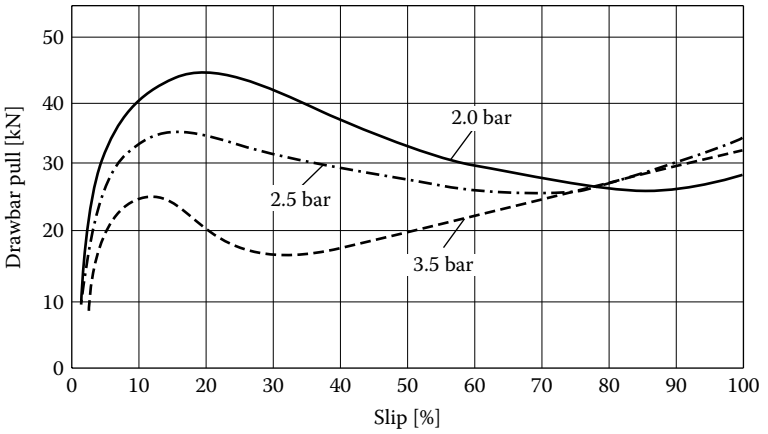


FIGURE 12.67 Influence of tire pressure on the DP–slip relationship. (Reproduced and modified from Schlechter, H.P., Design and performance of central tire inflation systems, *6th European ISTVS Conference*, Vienna, Austria, 1994.)

12.5.2.6.2.2 Auxiliary Supported Auxiliary-supported devices provide continued mobility in the event of partial or complete loss of tire air pressure [57–60]. Under combat conditions, these run-flat systems increase crew survivability by allowing them to complete the mission and return safely to base even if the vehicle's tires suffer complete air loss from hostile fire.

In this system, there is an additional support ring attached to the wheel that can support the weight of the vehicle in the event of a loss of pressure.

The inner tube shown in Figure 12.68 is an elastomer support ring in combination with a tubeless radial tire on airtight flat rims. Some other systems have steel rings to support the wheel load after puncture [60].

For tests, the puncture can be simulated either by perforating the tire by firing at it with rifle ammunition (five hits into the side wall and two into the tread) or by opening the valve.

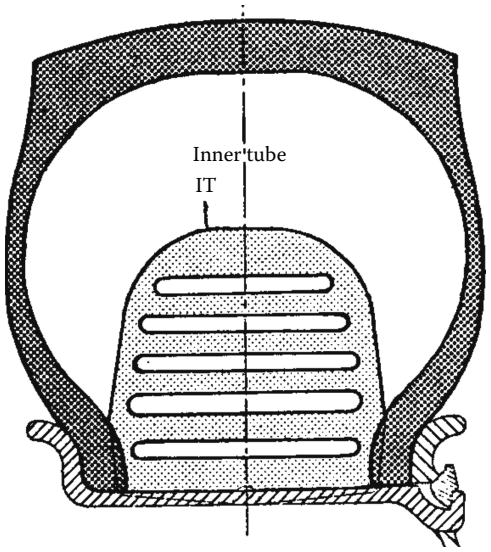


FIGURE 12.68 Auxiliary-supported run-flat system. (Reproduced from Hohl, G. and Pächter, H., Versuche an Reifen mit Notlaufeigenschaften, *FISITA Congress 1986*, Belgrade, Yugoslavia, 1986.)

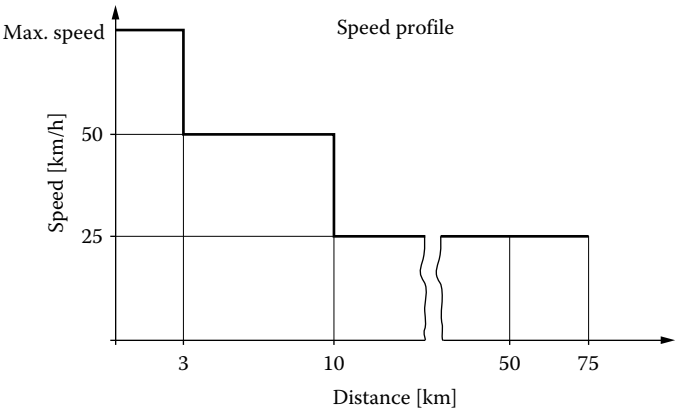


FIGURE 12.69 Speed profile for vehicle with puncture. (Reproduced from Hohl, G. and Pächter, H., Versuche an Reifen mit Notlaufeigenschaften, *FISITA Congress 1986*, Belgrade, Yugoslavia, 1986.)

Outline of the FINABEL Standard [61]

The vehicle with a puncture should be able to

1. Cover an overall distance of 50 km (preferably 75 km) with
   a. A tire still on the rim
   b. Without burning the carcass
2. Handling, steering ability, and speed of the vehicle should not be impaired excessively.
3. Distance should be covered according to the following speed profile (Figure 12.69):

The significant factors influencing the traction of a tire are *soil type* (see Section 12.3.1.1), *tire structure*, *tire diameter*, *tire width*, *wheel load*, *inflation pressure*, and *tread configuration*.

The radial tire is the most used type for passenger cars, trucks, and earthmoving machines. However, on the other hand, bias-ply tires are still in use for motorcycles, agricultural machines, and special military vehicles.

12.5.2.6.2.3 Tire Dimensions and Measurement Definitions The definitions of a tire are rather complicated, confusing, and in some cases not always logical. In this section, only principal definitions can be given. The variety of tire dimensions can be found in handbooks of tire manufacturers:

Nominal wheel diameter: Is the diameter of the rim seat supporting the tire bead. The bead-to-bead diameter is measured at the inside diameter of the tire, once mounted.
Section width: Is the maximum width of the cross section of the tire for an unloaded, mounted, and inflated tire.
Section height: Is the distance from the bead to the tread face.
Aspect ratio: Is the section height in relation to the section width, as percentage.

Tire definition:

The *first* number for a definition is the tire width (usually in inches, for smaller tires sometimes in mm).
The *second* number is the aspect ratio.
The *third* number is the nominal tire diameter (always in inches).
A radial tire is expressed by the letter R, set after the aspect ratio.

For example, a tire with the following dimension:

11/70 R 22.5: It is a radial tire and has a rim diameter of 22.5 in., a tire width of 11 in., and an aspect ratio of 70%.
Deflection δ: Is the difference between section height and loaded carcass section height.
Tread height: Is the inner tread shoulder to the outer tread shoulder.
Loaded section width: Is the section width under loaded conditions.
Tread width: Is the portion of the tread design that comes in contact with the ground.

Other measurements are

Overall diameter: Is the outside diameter of the tire when mounted and inflated but with no vehicle load.
Loaded radius: Is the distance from the wheel axle center line to the ground, drawn vertically. It is the distance from the wheel hub center line to the ground when the tire is inflated and supports vehicle weight (Figures 12.70 and 12.71).
Tire structure of terrain vehicles: An off-road vehicle requires modern radial tires with cross-country profile, which must be able to convert the output torque of the vehicle's engine into tractive effort on the ground. All off-road operations do not need the same degree of traction. As a result, separate tread designs are used for different degrees of tractive effort.

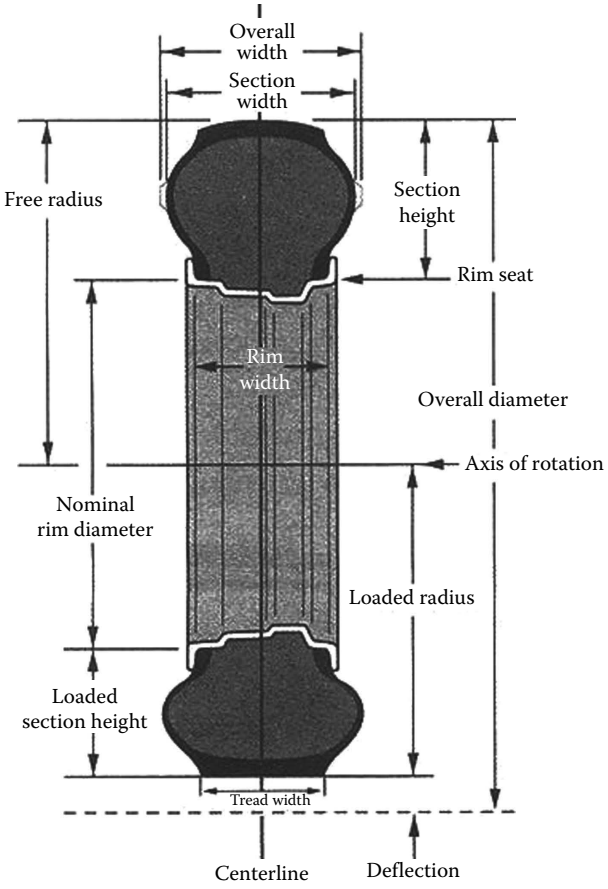


FIGURE 12.70 Tire dimensions. (Reproduced from http://www.actiontireco.com/Scripts/tiredefinitions.asp)



FIGURE 12.71 Conventional terrain tires. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

Off-road tires can be divided into two principal groups:

- Tires with the dimension of conventional tires but with special terrain treads
- Tires specially designed for off-road operations and with unusual dimensions (Figure 12.72)

Tire load: This influences its traction coefficient and hence its pulling capacity according to the type of soil. The traction coefficient decreases when the load on the tire is increased for loose or soft soils and increases for stiff types of soils. For certain types of soils (stiff frictional), the traction coefficient remains constant, or nearly constant, when the tire load changes for the same inflation pressure.

Tread configurations [62]: The requirements of a tread in terrain situations are to create high adhesive force in longitudinal and lateral directions with a minimum of slippage. Many efforts have been taken to predict a proper profile for wheeled terrain vehicles by computer.



FIGURE 12.72 Unconventional terrain tires. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

Tire treads with varying degrees of "aggressive" action have been tested to determine their direct influence on traction production. While no real analysis formulations have been derived to specifically quantify the degree of tread "aggressivity" in relation to traction production, much experience with tread design and geometry on tire performance has been gained in recent years through systematic tests.

In the construction field, as typified by dam, waterway, and highway projects that require movement of large quantities of earth and rock, tire types with high flotation characteristics and load carrying capacities are demanded.

The height of the lugs mainly affects the self-cleaning ability of the treads as well as the area of lateral ground shear on cohesive soils, which do not have an adjacent firm layer. With an increase in the height of the lugs, the area of ground shear and, consequently, the tire grip slightly increases.

The shape of the lugs influences the adhesion, self-cleaning ability, and flexibility of a tire. Due to the inclination of the supporting surface of the lugs, there is greater compaction of soils between the lugs, resulting in greater ground resistance to shear.

A *multipurpose tread* for off-road vehicles must provide for a high quality of performance, both on prepared automobile roads as well as on diverse types of soil- and snow-covered surfaces.

The principal requirements of an all-purpose tread are the absence of wheel vibrations while running on a hard surface. They should have low rolling losses on highways, good adhesion under diverse soil surface conditions, good self-cleaning ability, high structural strength, and durability under widely varying operating conditions (Figure 12.73).

The tread patterns of off-road tires are subdivided into "nondirectional" and "directional" types.

Nondirectional tread patterns

1. Combine a "checkerboard" pattern at the center of the tread band with long and sparsely dispersed transverse lugs at the sides. This pattern provides a comparatively high flexibility for the tread, smooth rolling over a hard top road, good adhesion on wet- and snow-covered roads, and a considerably higher adhesion compared to automobile tires. However, it is lower than that for off-road tires insofar as running on soft soils is concerned.
2. The "straight herringbone" tread pattern is less flexible compared to the previous one, but it provides extremely high adhesion on soft soils in a longitudinal direction. Adhesion in the transverse direction, however, is much lower. The tread has poor self-cleaning ability because of reduced flexibility of the lugs.
3. This pattern consists of triangular lugs separated from each other by longitudinal grooves. It provides high flexibility and better adhesion in the transverse direction.

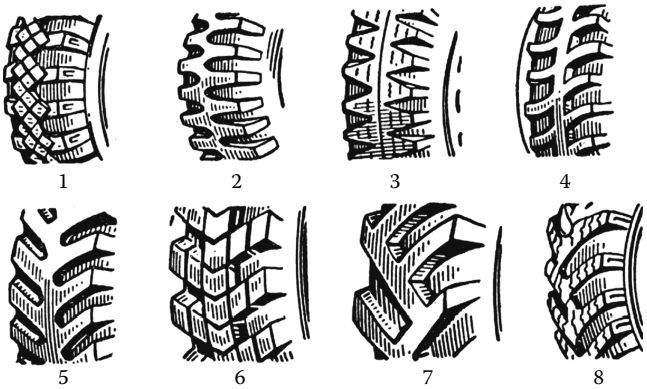


FIGURE 12.73 Tread patterns of "multipurpose" and "off-road" tires. (Reproduced from Ageikin, Ia.S., *Off-the Road Wheeled and Combined Traction Devices*, Oxonion Press PVT. Ltd., New Delhi, India.)

4. The "semi-spiral" pattern also has better adhesion in the transverse direction compared to the straight "herringbone" pattern tread. However, a lateral soil reaction component appears in the case of such a tread, which adversely affects the stability of the machine.

Directional tread patterns

5. The "herringbone with a continuous belt" at the center of the tread band is characterized by satisfactory self-cleaning and adhesive properties on soft soils. It ensures a relatively smooth ride over hard-covered roads. However, the tread is not sufficiently flexible.
6. The "disjointed herringbone" tread enhances the flexibility, reduces the rolling losses, and slightly improves the self-cleaning ability.
7. The "skew herringbone" tread pattern is similar to the tractor tread, except for a considerably larger area for the lugs with a smaller height. Such a tread pattern achieves better adhesion on soft soils compared to the herringbone pattern. However, for running on hard roads, it does not provide a smooth ride and, furthermore, the tread wear is uneven resulting in higher rolling losses.
8. The "disjointed skew herringbone" pattern is one of the most widely used tread patterns for off-road vehicle tires. The separation of the lugs by grooves increases the flexibility of the tread, reduces rolling losses, and improves the self-cleaning ability of the tire. Treads of this pattern may have either sharp or rounded shoulders.

Compared to the nondirectional treads, the directional tread pattern provides better self-cleaning ability and tire–soil adhesion over soft soils both in the longitudinal and transverse directions, when the tires are mounted in such a way that the vertex is pointed in the direction of rotation.

12.5.2.7 Tracks

Tracks have two basic functions:

1. To spread the load acting from the road wheels, to prevent sinking in soft soil
2. To transmit tractive effort generated by the engine

Tracks of armored and soft-skinned vehicles can be divided into

1. Band tracks
2. Segmented tracks
   a. One-pin track
   b. Two-pin track

Both types have a *rubber-coated* pattern or are equipped with *removable pads*. Now a new two-pin track is on the market, in which the sprocket has a toothing connection with the main body, like a one-pin track. This concept makes the track lighter and cheaper (Figure 12.74).

Band tracks are only in use for very small tracked vehicles like snow mobiles. However, there is now a trend to use this type also for heavier tracked vehicles up to 20 t. The limit for these endless one-piece molded tracks could be 30 t vehicles. Band track is a high-performance, low-maintenance alternative to the segmented metal track traditionally used for tracked combat vehicles.

A *segmented track* consists of pin-jointed steel elements.

In the case of a one-pin track, two chain elements are connected by a dry bearing pin. The sprocket drives the track via the steel casting main body. The disadvantage of this type is rattling noise and the short lifespan due to the dry bearing.

Several attempts have been made to improve the dry pin type of tracks like lubricated pins or needle bearings, but all these attempts were not successful.



FIGURE 12.74 Band track for an APC. (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

Nevertheless, one-pin tracks have the advantage of being cheaper and lighter in comparison with the two-pin track (Figures 12.75 through 12.77).

In two-pin tracks, the sprocket drives the track via the end connectors. Now a new two-pin track is on the market whereby the sprocket is in connection with the main body, like a one-pin track. This concept makes the track lighter and cheaper.

12.5.3 Suspension of Wheeled Terrain Vehicles

This section provides an overview of several suspension systems for wheeled off-road vehicles and a comparison between them. Conflicting suspension requirements for off-road purposes are reviewed [63–65].

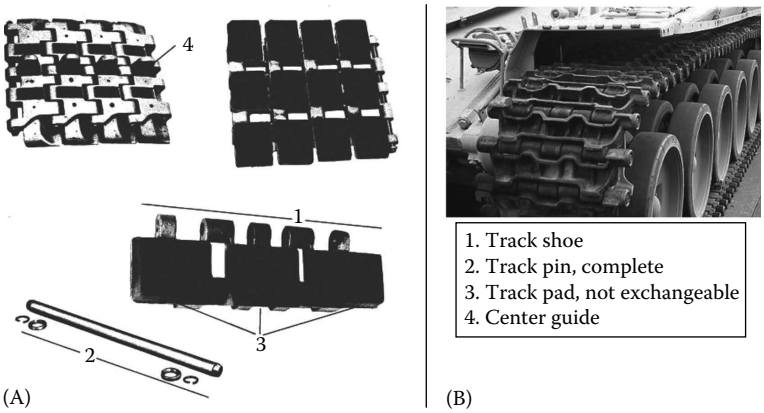


FIGURE 12.75 One-pin track with removable pads (A) and all steel track (B). (Reproduced from Folder Diehl Company, Modesto, CA.)

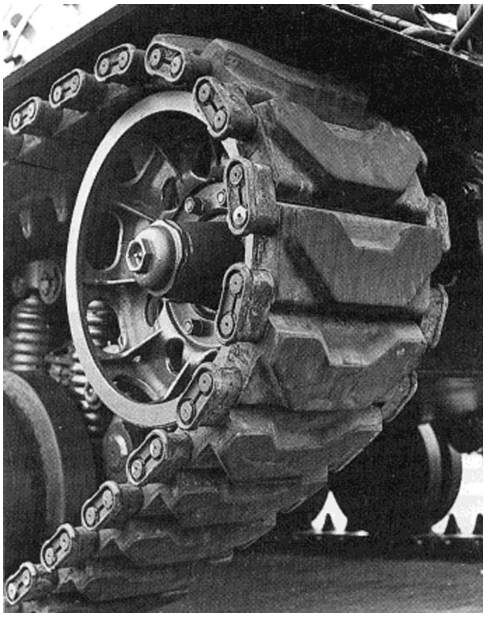


FIGURE 12.76 Rubber-coated track. (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

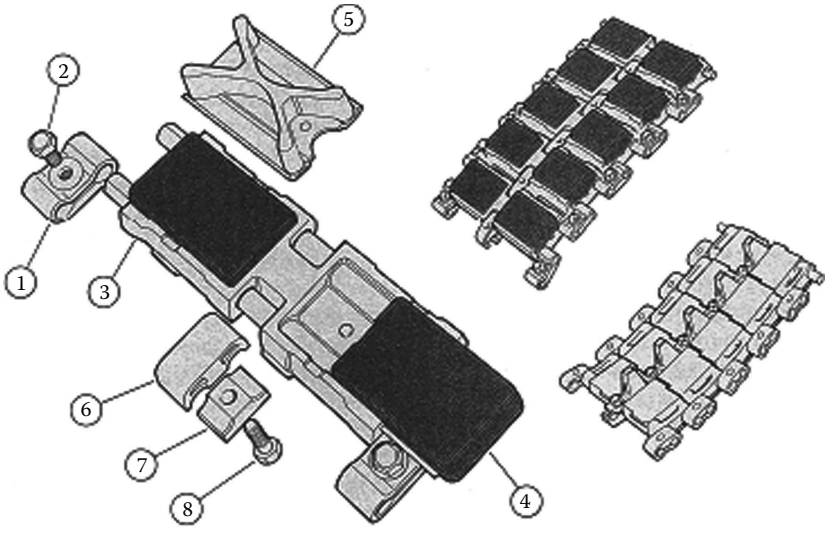


FIGURE 12.77 Diehl "system" track with removable pads. (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.). (1) End connector, (2) end connector bolt, (3) track link body with integral center guide, (4) replaceable rubber track pads, (5) grouser which can replace a track (for ice and snow conditions), (6) center connector, outer part, (7) center connector, inner part, and (8) center connector bolt.

The independent suspension is well established for passenger cars; however, the issue is not clear for trucks and off-road vehicles.

The mathematical models represented are based on the laws of mechanics.

12.5.3.1 General

The main consideration about suspension of wheeled vehicles is in many respects the same for armored and non-armored versions.

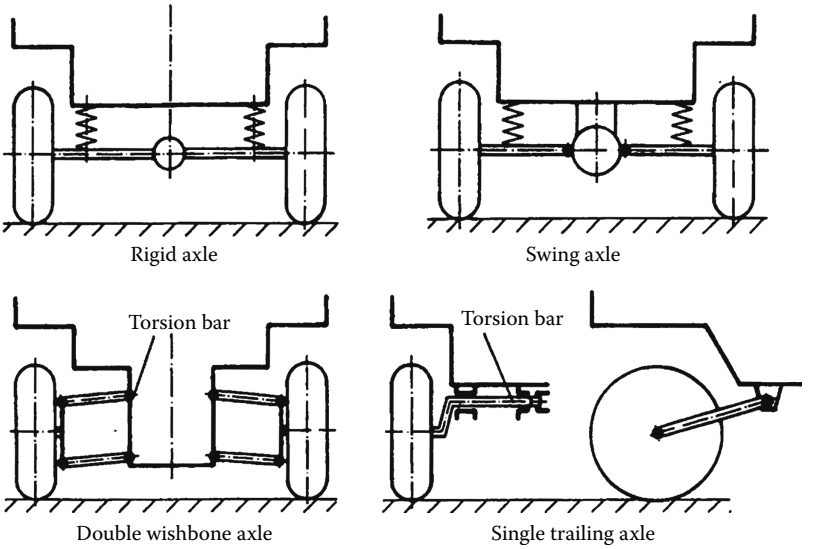


FIGURE 12.78 Main suspension systems for wheeled off-road vehicles. (Based on an idea from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, thesis, Technical University Vienna, Vienna, Austria, 1943; Reproduced from Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

According to the mechanical relationship between wheels, axles, springs, and the frame (hull), the suspension systems for off-road vehicles can be classified in two major categories, as in Figure 12.78:

- Non-independent types (rigid axle)
- Independent types (swing, double-wishbone, single trailing axles)

As both systems are being used successfully, it is worthwhile investigating their advantages and disadvantages. Suspension systems are examined with respect to

- Their ability to twist
- Body roll in curves and on slopes
- Wheel and steering control
- Economical, practical, and design considerations
- Overall height and configuration of vehicle bottom

The purpose of the suspension system of a wheeled vehicle is to support the weight of the vehicle, to separate the axles from the vehicle chassis, and thus to isolate the structure as far as possible from shock loading and vibrations caused by irregularities of the terrain.

Frames of modern off-road vehicles have a torsion-resistant chassis (see Section 12.5.1.1). Following this trend, the theoretical considerations of this section here are based on a rigid frame for soft-skinned vehicles and a rigid hull for AWV. When taking the flexibility of structures into consideration, methods and complicated computer models as FEM have to be used.

12.5.3.2 Non-Independent Suspension

Non-independent suspensions are characterized by rigid transverse structural members connecting the opposite wheels. With this arrangement, a vertical displacement of one wheel causes a displacement of the other wheel of the same axle.

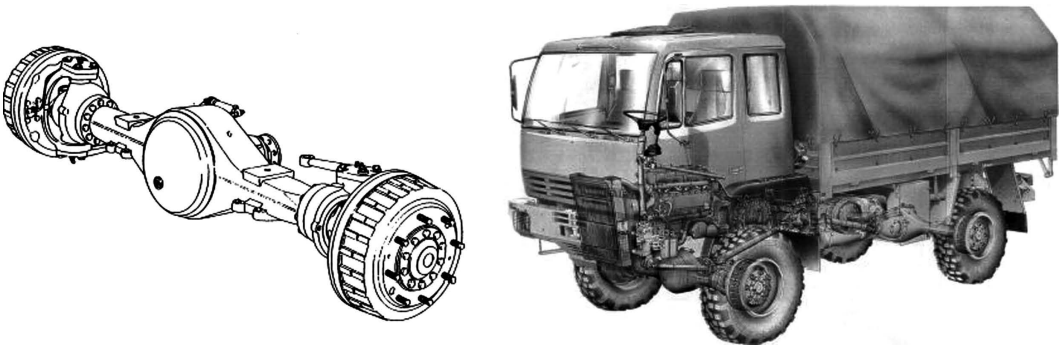


FIGURE 12.79 Medium military truck "Steyr 12M18." (Reproduced from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

The non-independent category consists only of one type, the *rigid axle*, sometimes also known as the *beam axle*, which is the oldest and most widely used system (Figure 12.79).

The rigid axle system with leaf springs is one of the most commonly used suspension systems for heavy commercial and military road and off-road vehicles. The torque reactions, driving thrust, and road-induced forces are transmitted to the springs that must be rigid enough to maintain reasonable wheel and axle alignment under these conditions.

Laminated springs consist of several layers, starting with the main leaf that is supported by shorter ones.

For road vehicles these semielliptic leaf springs are usually slung beneath the axles, but in off-road vehicles they rest on the axle to permit more ground clearance.

In the past years, a special development in the quality of leaf springs has taken place. New designs and new materials have been introduced. A recent design is the so-called taper leaf spring, also known as the parabolic leaf or minimum leaf spring. In this, the number of leaves is reduced and the thickness of the single leaf is increased (Figures 12.80 and 12.81).

Laminated springs consist of several layers, starting with the main leaf that is supported by shorter ones.

For road vehicles these semielliptic leaf springs are usually slung beneath the axles, but in off-road vehicles they rest on the axle to permit more ground clearance.

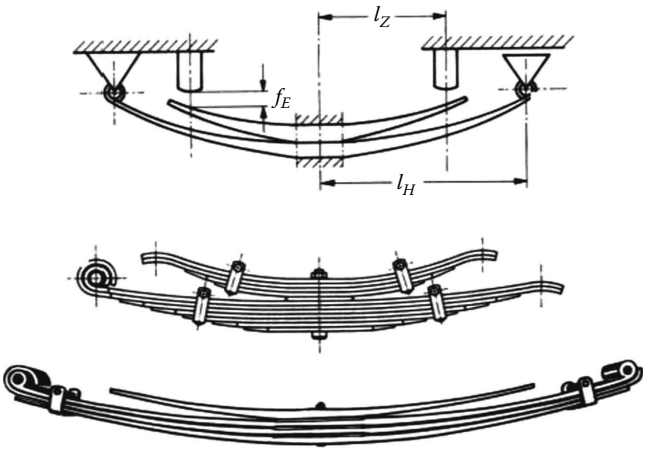


FIGURE 12.80 Leaf springs with a progressive spring characteristics. (Reproduced from Hohl, G. and Heimberg, N., The duro 4 × 4 and 6 × 6—A new concept for a family of soft-skinned and protected off-road vehicles, *EAEC 2001 Congress*, Bratislava, Slovak Republic.)

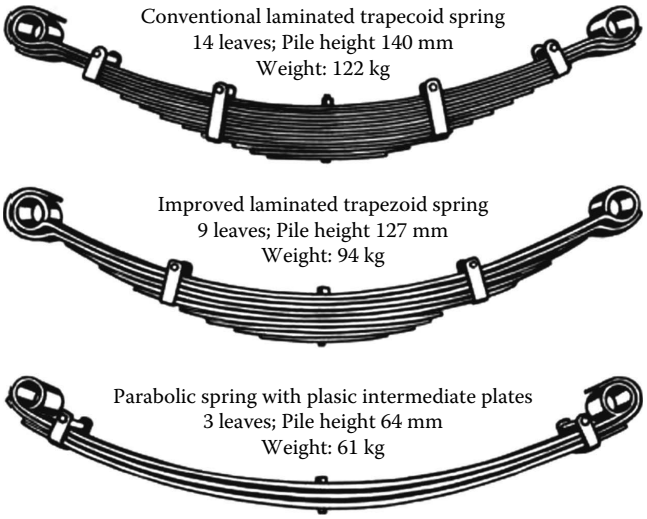


FIGURE 12.81 Leaf spring types with the same spring ability. (Reproduced and modified from Hohl, G. and Heimberg, N., The duro 4 × 4 and 6 × 6—A new concept for a family of soft-skinned and protected off-road vehicles, *EAEC 2001 Congress*, Bratislava, Slovak Republic.)

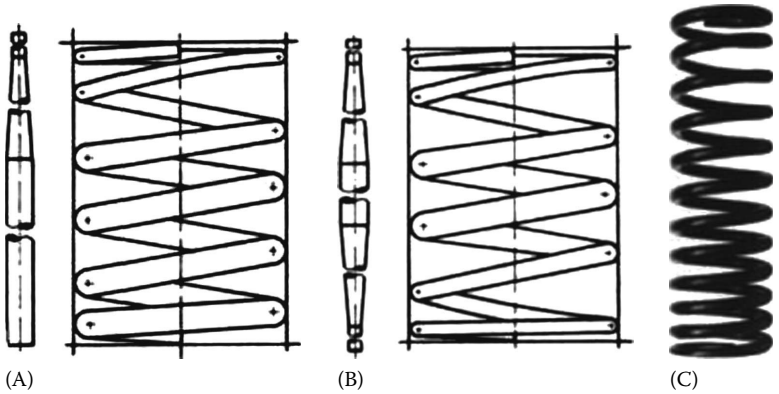


FIGURE 12.82 Coil spring types. (Reproduced and modified from Hohl, G. and Heimberg, N., The duro 4 × 4 and 6 × 6—A new concept for a family of soft-skinned and protected off-road vehicles, *EAEC 2001 Congress*, Bratislava, Slovak Republic.)

In the past years, a special development in the quality of leaf springs has taken place. New designs and new materials have been introduced. A recent design is the so-called taper leaf spring, also known as the parabolic leaf or minimum leaf spring. In this, the number of leaves is reduced and the thickness of the single leaf is increased (Figure 12.82).

The spring travel of leaf springs is limited. For extreme twisting ability and high spring travel, coil springs have to be used. One of the disadvantages with this type is the required volume for assembling. In addition, coil springs have no damping ability and no guiding ability for the axle.

Therefore, rigid axles equipped with coil springs need longitudinal control arms and transverse links (Figure 12.83).

12.5.3.3 Independent Suspension

Independent suspension is characterized by the lack of a rigid structural member to connect the individual wheels. With this arrangement, a vertical displacement of one wheel does not directly affect the other.

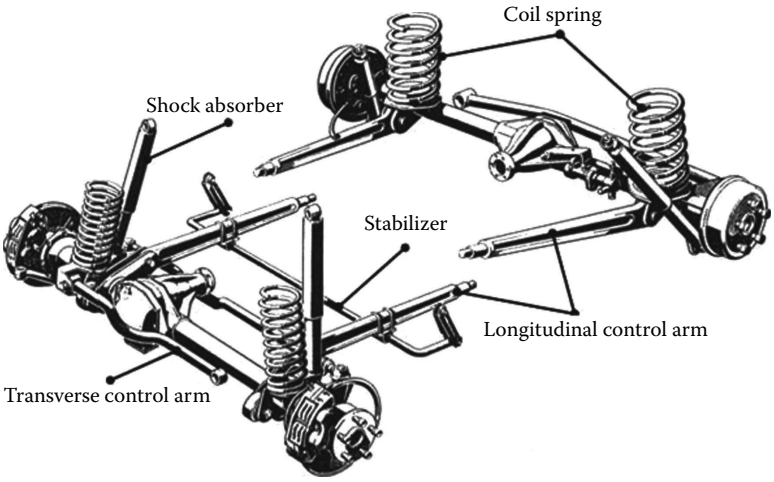


FIGURE 12.83 "Mercedes Puch G"; rigid axle with coil springs and longitudinal and transverse control levers. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

12.5.3.3.1 Swing Axle

A rarely used but very elegant design for independent suspensions is the swing axle, which is normally used in combination with a central tube frame. Two prominent representatives with a central tube frame and swing axles are the two legendary military off-road vehicles "Haflinger" and "Pinzgauer" and the heavy Czech truck "Tatra" (Figure 12.84).

The central tube reduces additional stresses and torsion of the body caused by wheel travel when riding over rough ground or negotiating an obstacle.

On the other hand, swing axles are a special design. Hardly any commercial components can be used. The connection between axles and differential is complicated and expensive. The ground clearance varies with loading. In the new "Pinzgauer," this problem has been solved by using a pneumatic lifting system to maintain constant ground clearance.

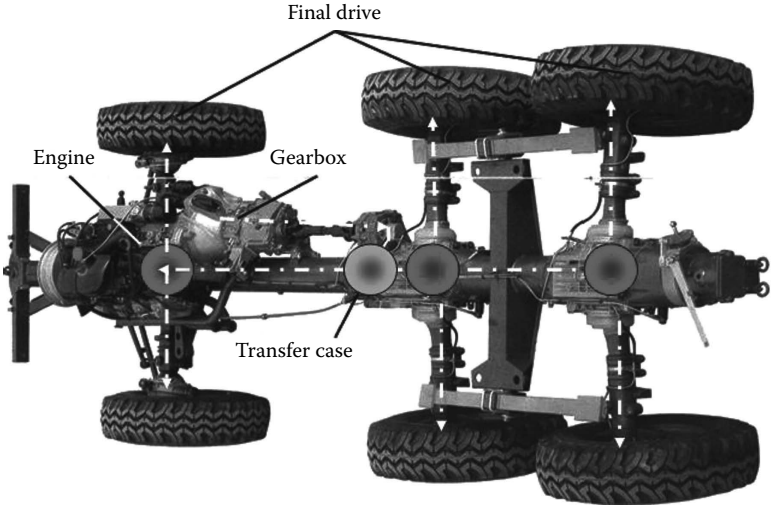


FIGURE 12.84 "Pinzgauer" 3-axle type. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

In addition to this, contrary to a ladder frame, the body can only be mounted on special points. This fact minimizes the variability of connecting the superstructures.

In the "Tatra" concept, this problem is solved with a subframe mounted on the central tube.

12.5.3.3.2 Double-Wishbone Axle

The double-wishbone axle is a very common concept for SUV and AWV.

The principal design for independent suspensions is the double-wishbone axle with two transverse links. If the two links are of the same length, there is an exact parallel guide of the wheel (Figure 12.85).

The springs can be coil springs or torsion bar springs that are mounted parallel to the longitudinal axis of the vehicle.

In history, there are some examples of independent suspension for off-road trucks but this concept is still a niche in the segment of medium and heavy trucks.

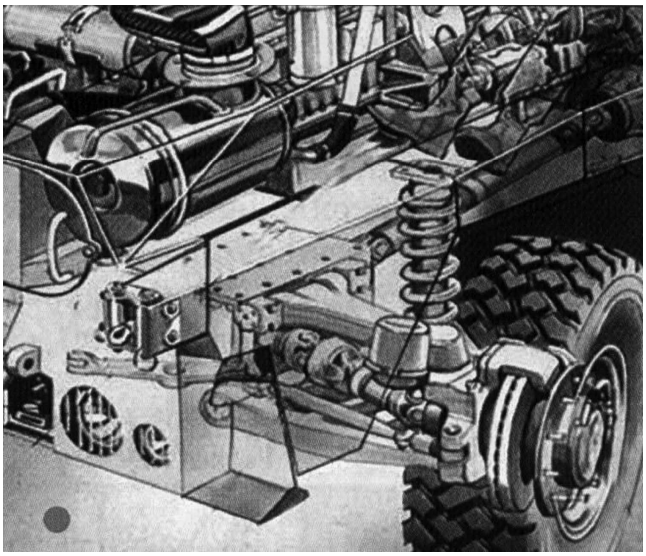


FIGURE 12.85 Double-wishbone axle of a soft-skinned military vehicle. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

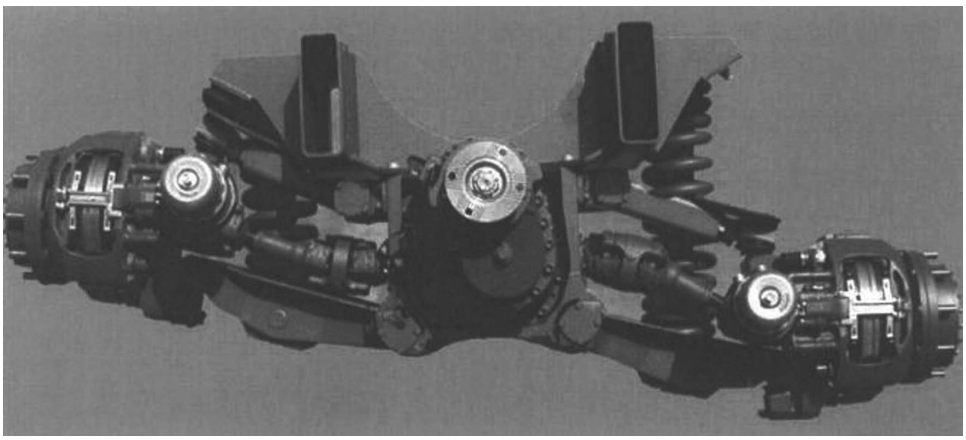


FIGURE 12.86 Double-wishbone axle concept for a heavy off-road truck; (system "Timoney"). (Reproduced from Folder of Timony Company.)

The Timoney Technology Group designs and develops independent suspension systems for a variety of vehicles for commercial and military applications, like APC, combat support vehicles, heavy transporters, and airport crash fire rescue vehicles (Figure 12.86).

12.5.3.3.3 Single Trailing Axle

The single leading or single trailing suspension has its origin in the road wheel design of tracked vehicles. The spring is usually a torsion bar that is mounted crosswise to the longitudinal axis. Therefore, trailer axles are mainly used for non-steered axles or for skid-steering wheeled vehicles.

However, there are also exceptions. The agriculture vehicle series "Reform Muli" has a steerable front axle with trailing axle principle and coil spring system.

An example of this principle is explained in Figure 12.87, the Austrian wheeled AVC "Pandur," which has a combination of a modified double-wishbone axle on the first two axles and a single trailing axle in the rear.

12.5.3.3.4 Unusual Suspension Concepts

Beside the four basic and conventional axle design principles, there are some more interesting solutions on the market (Figures 12.88 and 12.89).

12.5.3.3.5 Active Suspension

The primary function of suspension systems for off-road vehicles is to isolate the crew compartment vibrations caused by terrain irregularities and to maintain running gear contact with the ground. A conventional passive suspension system accomplishes this by using springs and constant damping rates [66–71].

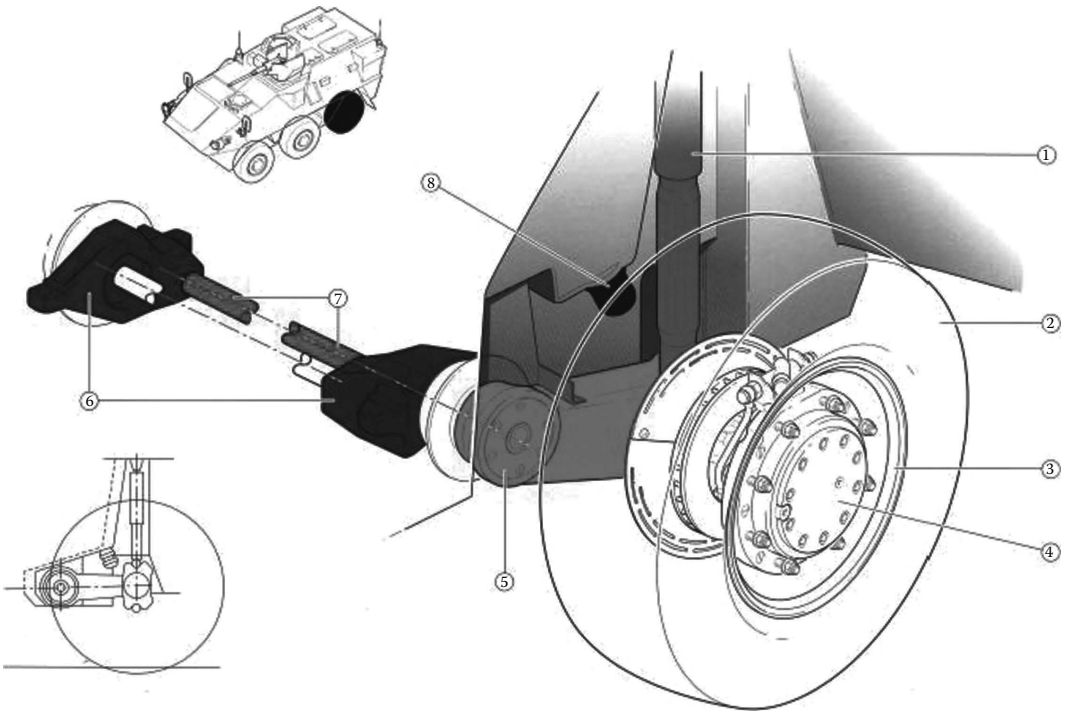


FIGURE 12.87 Single trailing axle (3 axle) of the Austrian AVC "Pandur." (Reproduced from Folder of Steyr Spezialfahrzeuge Company.)

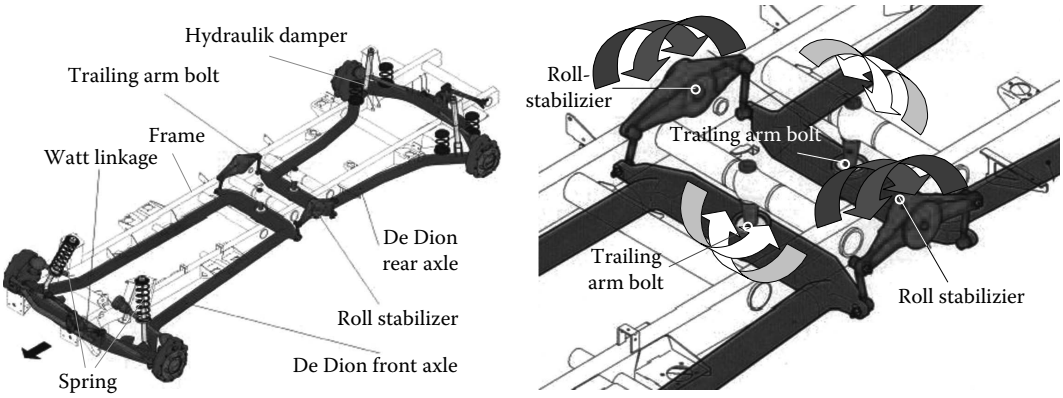


FIGURE 12.88 Swiss light military vehicle "MOWAG DURO." (Reproduced and modified from Nell, S., *J. Terramech*., 31(4), 227.)

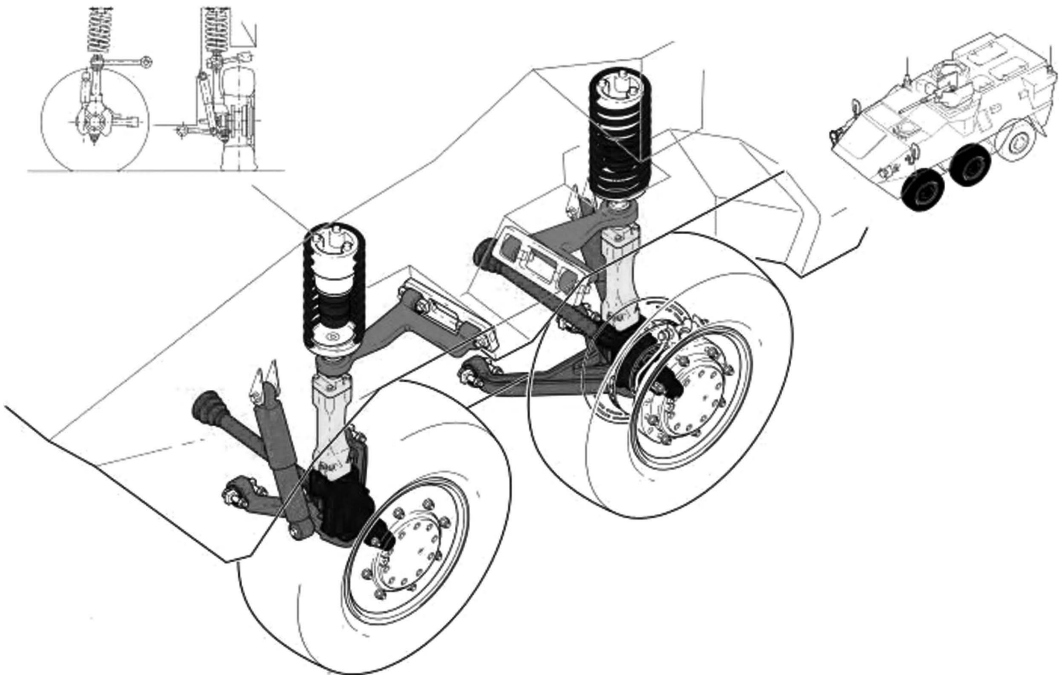


FIGURE 12.89 Modified double-wishbone axle (1 and 2 axle) of the Austrian wheeled AVC "Pandur." (Reproduced and modified from Folder of Steyr Spezialfahrzeuge Company.)

12.5.3.4 Twisting Ability

12.5.3.4.1 General

The differences between the wheel loads ΔW_i of diagonally opposed wheels while negotiating an obstacle can be regarded as one of the indicators of off-road mobility. Equal wheel loads result in more uniform tire adhesion for the production of thrust. In the case of unequal wheel loads, the wheels with the higher load cause increased rolling resistance. The other wheels with decreasing load cause a decrease of traction.

The model used for the mathematical equations is described in Figure 12.90.

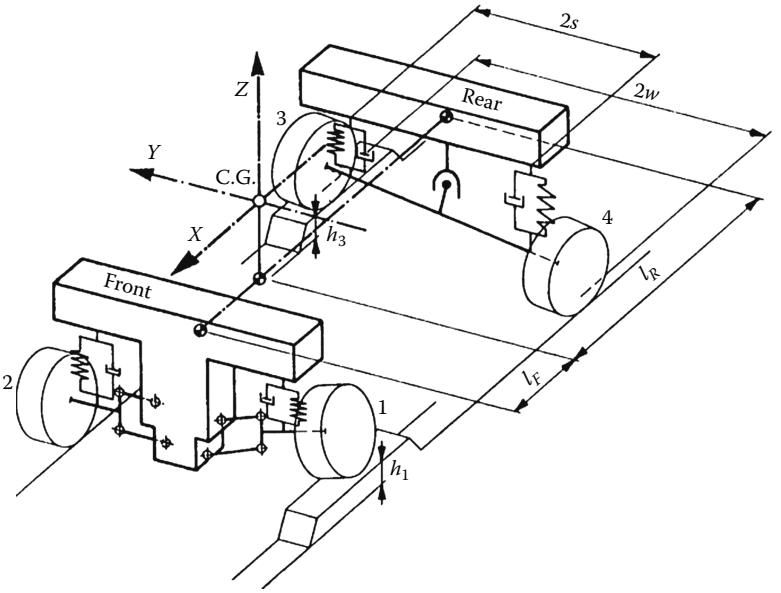


FIGURE 12.90 Basic model for twisting ability. (Reproduced from Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

Obstacles are put in front of two diagonally opposed wheels—in this figure, the left front wheel and the right rear wheel. Then the heights of the obstacles are increased until an opposite wheel loses ground contact. The purpose of the mathematical investigations is to find a correlation between obstacle heights h_1 and h_3 and wheel load difference ΔW_i for several suspension systems.

Twisting ability h is characterized by the sum of the height (h_1 and h_3) of the obstacles.

$$h = h_1 + h_3 \tag{12.43}$$

The wheel load difference ΔW_i between before and after twisting is defined as follows:

$$\Delta W_i = W_{i0} - W_i \tag{12.44}$$

where

i is the wheel index (1–4)

0 is the situation before twisting that means the vehicle stands on a horizontal plane

The maximum twisting ability is reached when a wheel lifts from the ground.

$$\Delta W_i = W_{i0} \tag{12.45}$$

To make the different axle designs comparable, the spring characteristic is expressed by the travel of the vehicle body (φ_i) under normal load.

For the determination of φ_i for the different axle systems, the geometrical relationships of the axle design and the springs have to been taken into consideration.

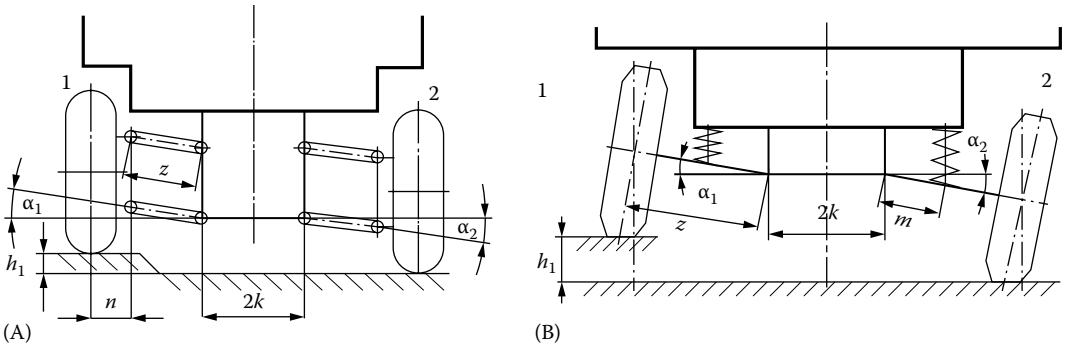


FIGURE 12.91 Double-wishbone axle (A); swing axle (B). (Reproduced from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, thesis, Technical University Vienna, Vienna, Austria, 1943; Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

The following equations are valid for all types of axles:

- The sum of the wheel loads are equal to the total load.

$$\sum_{i=1}^{4} W_i = W \tag{12.46}$$

- The sum of the momenta through the *X*- and the *Y*-axis is zero.
  (Only the torque through the *Y*-axis can be expressed as one equation for all suspension systems.

$$(W_1 + W_2)l_F = (W_3 + W_4)l_R \tag{12.47}$$

- The equations for the other axle systems depend on the suspension types and can be derived from the geometrical relationships in Figure 12.91.

In order to make the different axle designs comparable, the spring characteristic φ is expressed by the wheel travel under normal load, when the vehicle body is regarded as a rigid basis. A more detailed derivation of the equations can be found in Ref. [63].

12.5.3.4.2 Rigid Axle

In short, only the mathematical derivation of the equations for the rigid axle is given. First, the load of the four wheels ($W_1 \ldots W_4$) and the angles α_F and α_R are unknown.

To determine the six unknown values, six equations are available (Figure 12.92).

Rigid Axle

$$\frac{\Delta W_1}{W} = -\frac{\Delta W_2}{W} = h\frac{l_F l_R s_F^2 s_R^2}{4w_F(l_F + l_R)(l_F s_R^2 w_F \varphi_F + l_R s_F^2 w_R \varphi_R)} \tag{12.48}$$

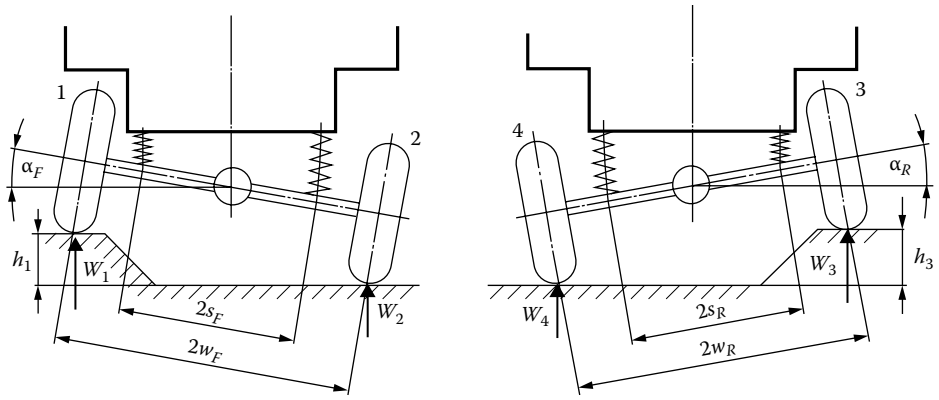


FIGURE 12.92 Rigid axle while negotiating diagonal obstacles. (Reproduced from Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

$$\frac{\Delta W_3}{W} = -\frac{\Delta W_4}{W} = h\frac{l_F l_R s_R^2 s_F^2}{4w_R(l_R + l_F)(l_R s_F^2 w_R \varphi_R + l_F s_R^2 w_F \varphi_F)} \tag{12.49}$$

When the wheel, respectively, spring tracks w_F, w_R and s_F, s_R as well as φ_F and φ_R are the same for both axles, the equation can be extremely simplified. A further simplification is to put the C.G. in the middle between the axles:

$$\frac{\Delta W_i}{W} = \frac{h}{16\varphi}\left(\frac{s}{w}\right)^2 \tag{12.50}$$

According to Equation 12.45, the maximal obstacle heights for the front and the rear axle can be determined as

$$h_{max(F)} = \frac{2w_F(l_F s_R^2 w_F \varphi_F + l_R s_F^2 w_R \varphi_R)}{l_F s_F^2 s_R^2} \tag{12.51}$$

$$h_{max(R)} = \frac{2w_R(l_F s_R^2 w_F \varphi_F + l_R s_F^2 w_R \varphi_R)}{l_R s_F^2 s_R^2} \tag{12.52}$$

The smaller value of the two equations leads to the maximal twisting ability.

Simplifying these equations, in the same way mentioned earlier, derives to the following relationship:

$$h_{max} = \frac{4\varphi}{(s/w)^2} \tag{12.53}$$

It can be clearly seen that the maximum obstacle height increases with high φ-values (soft spring) and decreases with increasing (*s*/*w*) ratios (the spring is near to the tires).

12.5.3.4.3 Independent Suspensions

The wheel load differences of wishbone and swing axle are independent of the wheel track and spring track relationship. By setting for small values $\sin\alpha = \alpha$ and $\cos\alpha = 1$, the same equation for all three types of independent suspension can be determined:

$$\frac{\Delta W_1}{W} = -\frac{\Delta W_2}{W} = h\frac{l_R l_F z_F w_R}{4(l_F + l_R)(l_F w_R z_F \varphi_F + l_R w_F z_R \varphi_R)} \tag{12.54}$$

$$\frac{\Delta W_3}{W} = -\frac{\Delta W_4}{W} = h\frac{l_F l_R z_R w_F}{4(l_F + l_R)(l_R w_F z_R \varphi_R + l_F w_R z_F \varphi_F)} \tag{12.55}$$

With the simplifications mentioned earlier, the maximal obstacle height can be determined as follows (Figure 12.93):

$$h_{max} = 4\varphi \tag{12.56}$$

The specific wheel-load difference ($\Delta W_i/W$) is less for the rigid axle than for independent suspension systems. The wheel-load difference for the rigid axle depends on the relationship between wheel and spring track. In this diagram, the spring characteristic is equal for both front and rear axle. The maximum height (h) of an obstacle is achieved when the wheel opposite to the wheel on the obstacle loses ground contact, which occurs when the wheel load is zero.

The wheel-load differences for this kind of suspension are independent of the wheel track and spring track relationship. Usually spring track (s) and wheel track (w) are equal for both axles and the equation is easier to understand by looking at the s/w relationship (Figure 12.94).

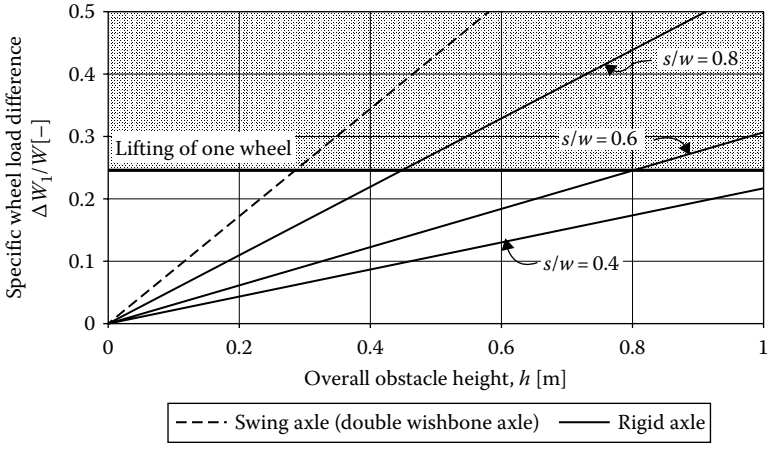


FIGURE 12.93 Specific wheel load difference versus obstacle height. (Reproduced from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, thesis, Technical University Vienna, Vienna, Austria, 1943; Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

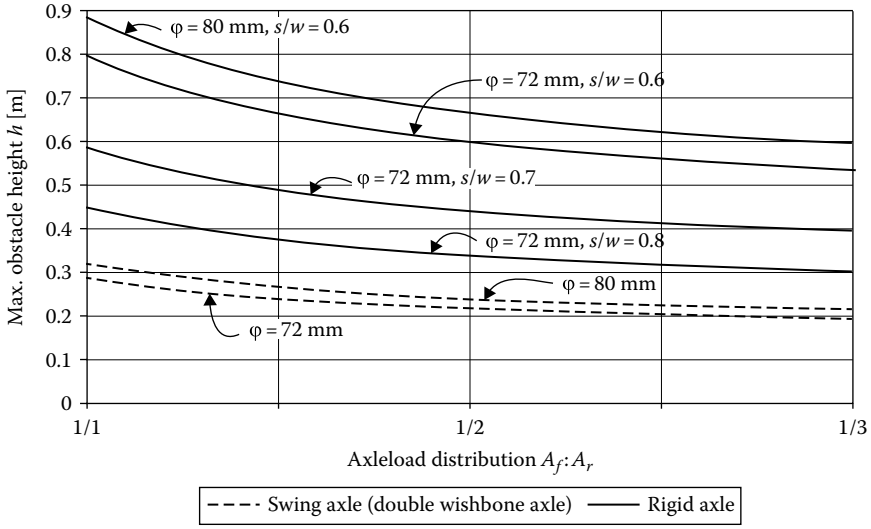


FIGURE 12.94 Max. obstacle height versus axle load distribution, spring stiffness, and axle concept. (Reproduced from Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

12.5.3.5 Body Roll

12.5.3.5.1 General

Off-road vehicle body roll in curves and on slopes is also an important indicator for cross-country ability. The roll angle should be small. A simple method to achieve this would be to use stiff springs. However, for reasons of ride comfort and to protect both the vehicle itself and the cargo, the use of springs with stiff characteristics is not practical. Curve and slope stability are influenced by the height of the roll center. The roll center is the apparent stationary point around which the body rolls under the influence of centrifugal forces and slope components. The distance between the roll center and the C.G. is regarded as a lever for the torque exerted by roll forces and has a great influence on the roll angle.

The centrifugal force C is defined as follows (R is the radius of the curve, v is the vehicle speed, g is the gravitational acceleration) (Figure 12.95):

$$C = \frac{W}{g} \cdot \frac{v^2}{R} \tag{12.57}$$

12.5.3.5.2 Rigid Axle

The roll center 0 of a rigid axle is easy to determine and it is located in the middle of the axles and the springs. With respect to different axle loads and different wheel and spring distances, it causes a very complicated equation. Making the simplifications mentioned earlier, the following equation can be derived:

$$\alpha_c = \frac{C \cdot (h - h_f)}{(s^2/\varphi) - (h - h_f)} \tag{12.58}$$

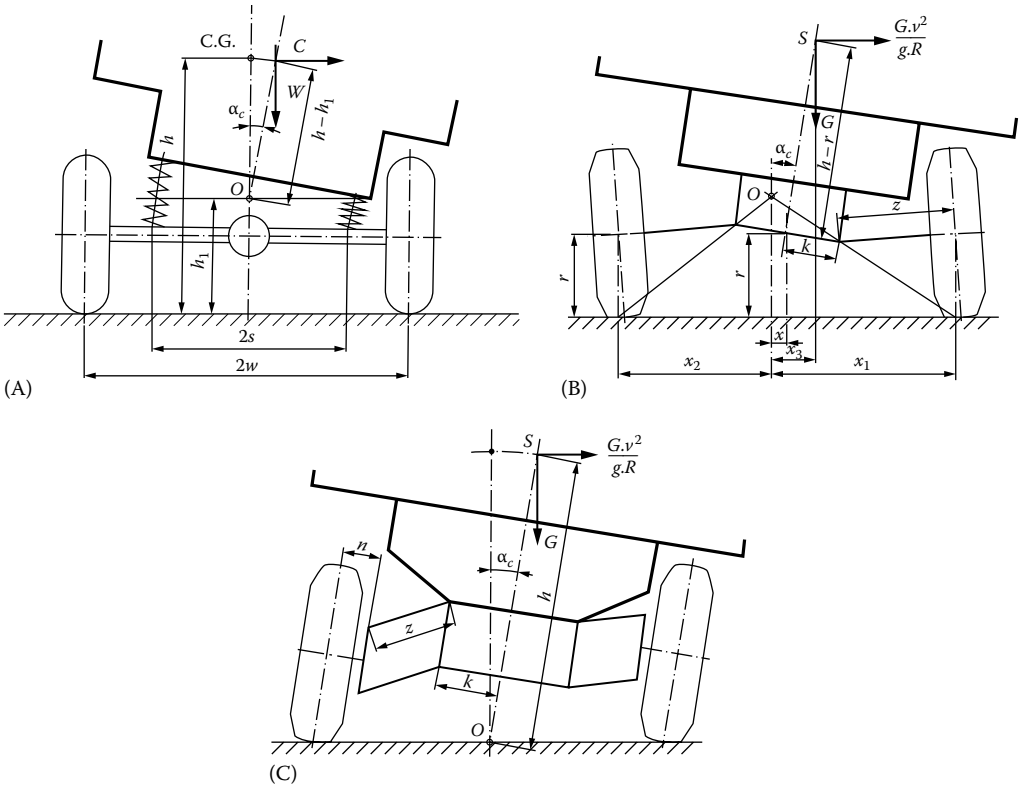


FIGURE 12.95 Centrifugal forces acting on a rigid axle (A) on a swing axle (B) and on a wishbone axle (C). (Reproduced from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, thesis, Technical University Vienna, Vienna, Austria, 1943; Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

12.5.3.5.3 Independent Suspensions

The roll angle α_c is different for swing and double-wishbone axles:

Swing Axle

$$\alpha_c = C\frac{h-(w/z)\cdot r}{(w^2/\varphi)-(h-r+(w/r)\cdot k)} \tag{12.59}$$

Double-Wishbone and Swing Axle

$$\alpha_c = C\frac{h}{(w^2/\varphi)-h} \tag{12.60}$$

In Figure 12.96, the roll angle versus the several suspension systems is shown. The best system is the one with the lowest roll angle. The rigid axle has lower roll stability. The best situation can be seen in the swing axle. The roll angle for the double-wishbone axle is independent of the geometrical relationship.

Figures 12.91 and 12.92 show the influence of the spring characteristic and the geometry of the axles on the roll angle for several suspension systems.

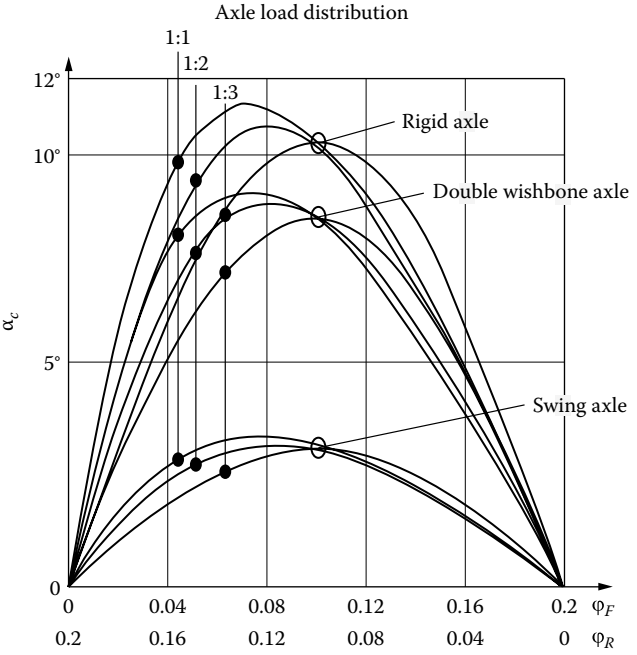


FIGURE 12.96 Roll angle versus suspension system considering different axle load distributions and different spring rates. (Reproduced from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, Thesis, Technical University Vienna, Vienna, Austria, 1943; Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

12.5.3.6 Steering Ability and Wheel Control

Passing over terrain irregularities and negotiating obstacles is one major task for off-road vehicles. When one wheel of a rigid axle suspension passes over an obstacle, the axle executes angular movement. The wheels are simultaneously displaced by approximately equal and opposite amounts relative to the chassis. Since wheels in movement act as gyroscopes, a forced gyroscopic movement occurs that tends to make the steerable wheels swing about their kingpins. This effect can be seen clearly in Figure 12.97 without any mathematical considerations.

Another problem associated with rigid axle design is that the unsprung masses are higher than for any independent suspension system. This is undesirable when both wheels are crossing an obstacle at the same time because the mass of the whole axle is involved.

12.5.3.7 Further Considerations and Summary

Ground clearance is a very important factor in off-road mobility. The rigid axle design has a constant ground clearance. The lowest point of the suspension system is the differential housing. In contrast to the independent suspension systems, the ground clearance is constant and independent of loading. On the other hand, the differential housing is an obstacle when passing high terrain irregularities. In the case of an independent suspension, the vehicle's belly can be designed to be flat and the clearance is not obstructed by axle beams.

Another disadvantage of a rigid axle system is the overall height. This difference can be seen clearly in an example for a wheeled armored vehicle. The rigid axle requires clearance between the axle and the hull to permit spring action (Figure 12.98).

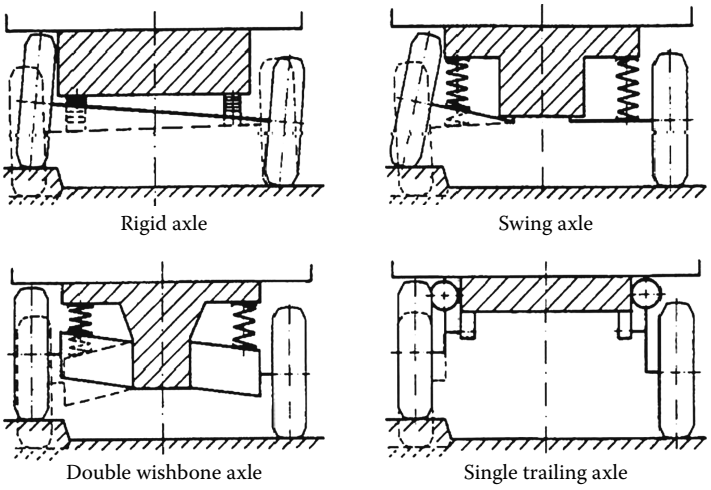


FIGURE 12.97 Wheel control for the main suspension types. (Reproduced from Klanner, R., Die Radaufhängung des geländegängigen Kraftfahrzeuges, Thesis, Technical University Vienna, Vienna, Austria, 1943; Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

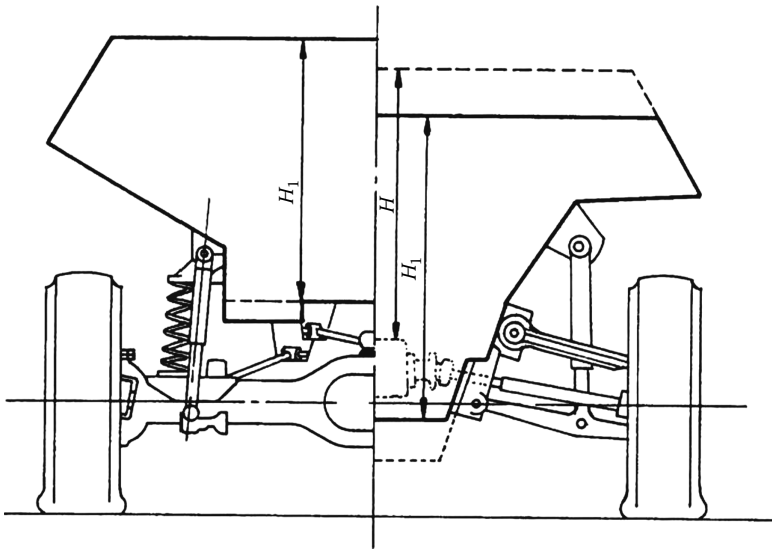


FIGURE 12.98 Wheeled armored vehicle with non-independent and independent suspension designs. (Reproduced from Hohl, G.H., The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993; Hohl, G.H., An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.)

12.5.4 Suspensions of Tracked Vehicles

12.5.4.1 General

The first tanks had nothing between their hull frames and the tracks other than rigidly mounted rollers. This was acceptable only for operations at low speeds for which these vehicles were designed.

In the first decades of the introduction of tracked armored vehicles, all kinds of mechanical spring systems were in use. Leaf, truncate, and disc springs vanished.

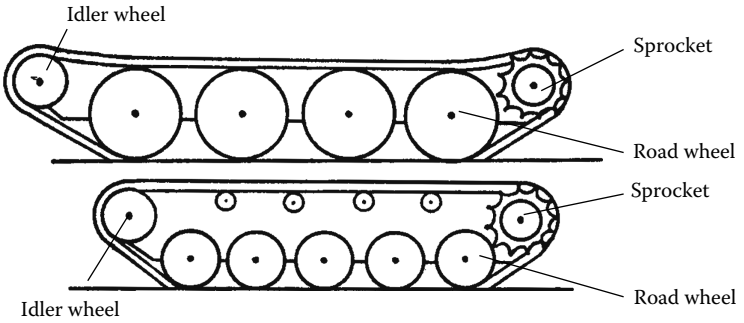


FIGURE 12.99 "Christie" suspension and supporting roller suspension. (Reproduced and modified from Hohl, G.H., Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.)

Coil springs in some niches are still in use. The torsion bar system is the most used spring on tracked armored vehicles but the number of tanks with hydropneumatic suspension is increasing. The most innovative spring system is the hydropneumatic suspension.

For the suspension system of tracked vehicles, two principal systems are in use:

- "Christie" or road wheel suspension
- Supporting roller suspension (Figure 12.99)

The "Christie" suspension supports the upper part of the track by the road wheels. Contrary to this in the supporting roller system, this function is done by the support rollers. Both systems have their advantages and disadvantages.

The rolling resistance decreases with the number diameter, but the diameter is limited to 800 mm. On the other hand, smaller and thus more road wheels spread the weight of the vehicle evenly over the length of the track weight that reduces the peaks of pressure on the ground under the road wheels. The use of support rollers to guide the return run of the track provides larger road wheel travel in vertical direction, causes lower unsprung masses, and avoids either excessive sag of the run of the track, which leads to the track being thrown off, or excessive track tension, which increases rolling resistance. The "Christie" suspension is used for tracked armored vehicles with a lower demand of terrain mobility like self-propelled howitzers.

12.5.4.2 Torsion Bar Suspension

Torsion bar suspension uses the flexibility of a steel bar twisting lengthwise to provide spring actions. One bar of spring steel is used for each road wheel in this type of suspension. One end of the bar is fixed solidly on the opposite side of the hull. The arm rises and falls with the road wheel movement [72] (Figure 12.100).

The main reason for the widespread adoption of a torsion bar suspension has been its light weight and the high levels of performance. The light weight is due to the basic simplicity and it stores more energy in relation to its weight than other spring systems.

The advantages of torsion bar suspensions are, however, accompanied by a number of disadvantages. Thus, while the installation of torsion bars across the bottom of tank hulls is simple and well protected, it also increases the height of the armored vehicle (Figure 12.101).

During WW II, torsion bar suspensions were still being designed to a maximum shear stress of 300 MN/m^2, but the improvements in the manufacturing of spring steels resulted in a maximum shear stress of 800 MN/m^2 in the fifties.

Advances in metallurgical techniques increase the maximum shear stress still further. Thus, modern torsion bar steels are able to operate at up to 1240 MN/m^2 for 68,000 cycles of loading.

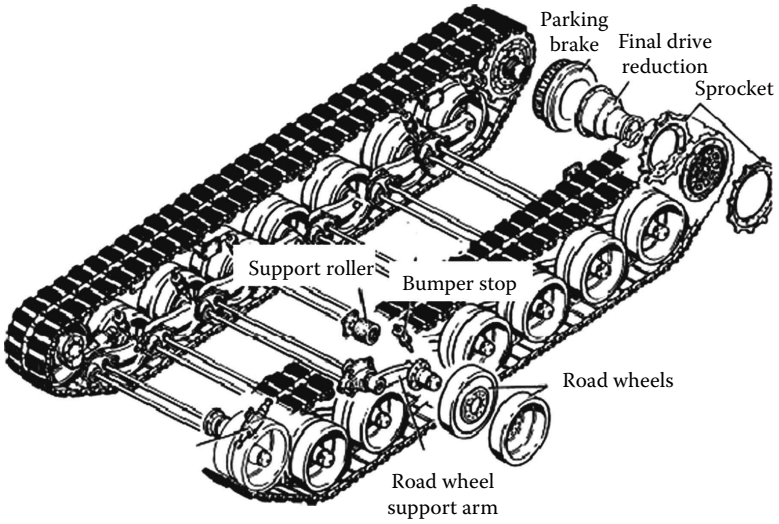


FIGURE 12.100 "Torsion bar" and "support roller" suspension of the German MBT "Leopard 2." (Reproduced and modified from Spielberger, W.J., *Von der Zugmschine zum Leopard 2*, Bernhard & Graefe Verlag, München, Germany.)

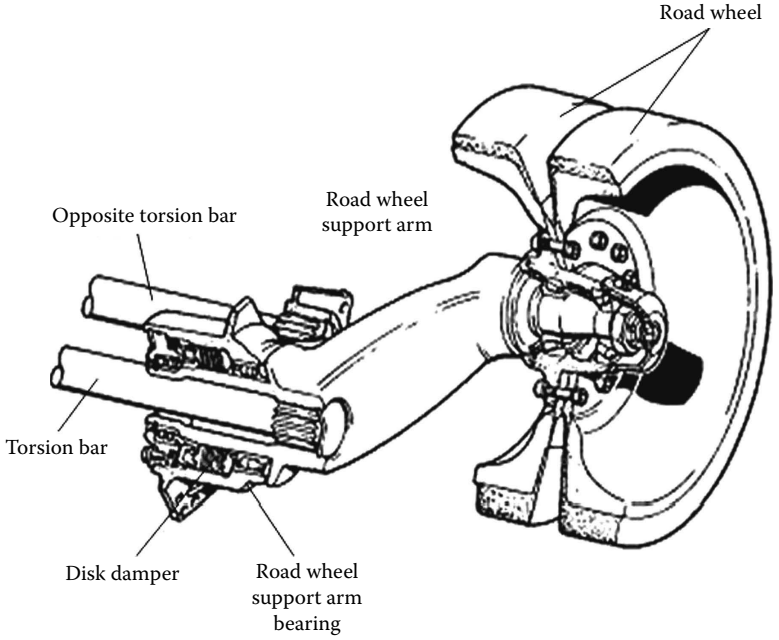


FIGURE 12.101 Torsion bar arrangement. (Reproduced from Spielberger, W.J., *Von der Zugmschine zum Leopard 2*, Bernhard & Graefe Verlag, München, Germany.)

12.5.4.3 Coil Springs

Coil springs have an old tradition. This kind of suspension was commonly used between the two world wars. Examples are the British "Cromwell" tank and the Soviet "T34."

The Israeli MBT "Merkava" has an independent suspension that means each road wheel has an independent spring assembly.

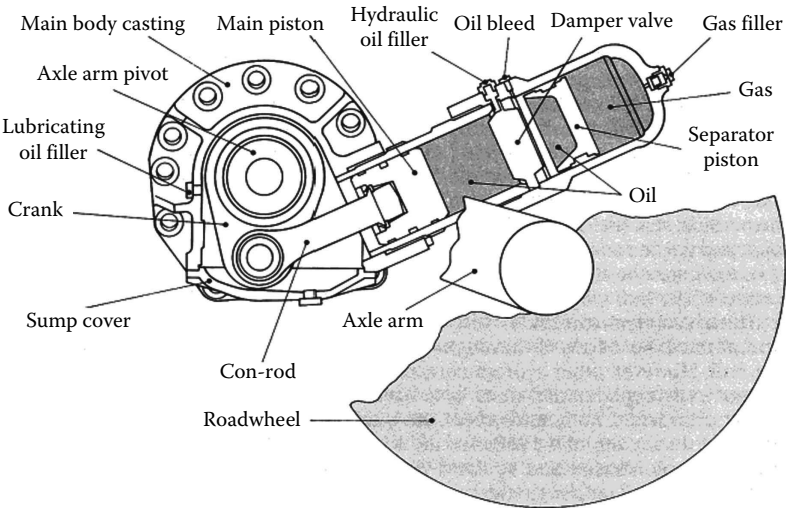


FIGURE 12.102 Hydropneumatic suspension. (Reproduced from Ogorkiewcz, R.M., *Tank Technology*, Jane's Information Group, London, U.K.)

12.5.4.4 Hydropneumatic Suspension

The hydropneumatic suspension contains a volume of gas, separated from a floating piston, or a diaphragm, from hydraulic fluid that transmits pressure between it and a second piston connected to the road wheel arm. Their resilience is based on the compressibility of the gas. Hydropneumatic spring units have progressive load–deflection characteristics.

Each unit is a spring as well as a damper. They provide better damping than suspensions in which dampers are only fitted to some of the road wheels. Moreover, when they are bolted to the hull, they can better dissipate the generated energy by telescopic dampers because the hull constitutes a large heat sink (Figure 12.102).

12.5.4.5 Dampers

There are two principal dampers in use, the *linear* and *rotary* dampers. The linear dampers are well known in private cars and in commercial vehicles. These dampers have the same principle but a greater size. The disadvantage is that mud can cling to the space between the damper and the other parts of the suspension. Rotary dampers have the same principle, but the motion of the piston is a rotary one.

For torsion bar suspensions, these kinds of dampers save space and are not the source of accumulation of mud. Rotary dampers can be fitted in the bearing of the torsion bar or connected with a lever to the rod arm of the road wheel (lever actuated).

As resonance can result in large oscillations, suspensions need to be well damped, as already indicated. On the other hand, damping should be as low as possible to minimize the transmission of impact forces, especially at high speeds or, what amounts to much the same thing, at high forcing frequencies.

In most cases, dampers have been fitted only to the front and rear road wheels. The reason for not fitting them to the center wheels is that they are ineffective there in damping pitching resonance, which sets in at lower frequencies than bounce resonance and is generally more disturbing. In fact, bounce resonance not only sets in later than pitch resonance but does so at frequencies beyond the cross-country speed of many vehicles. However, damping of the vertical oscillations becomes important at high speeds.

Friction dampers have also been used, their principal attraction being greater durability than that of hydraulic dampers. On the other hand, friction damping has been less consistent and has not provided as good a range of characteristics as hydraulic damping. Nevertheless, highly developed, disc-type, rotary friction dampers have been adopted for the "Leopard 2" tank (Figure 12.103).

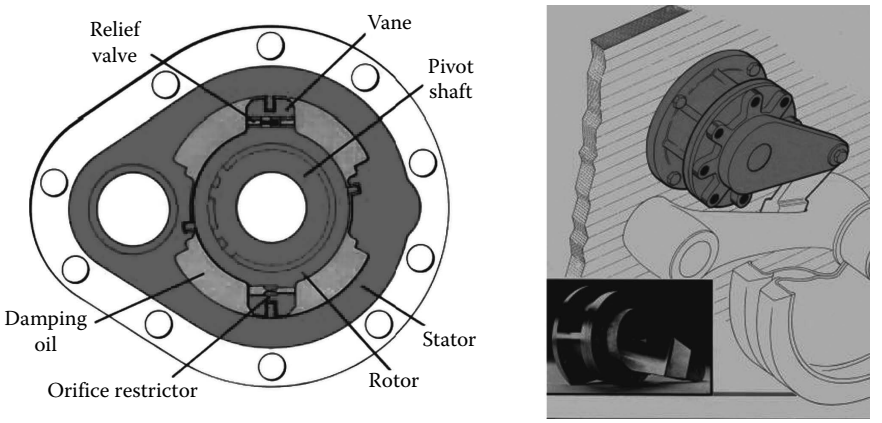


FIGURE 12.103 Rotary dampers. (Reproduced from Holman, T.J., Rotary hydraulic suspension dampers for high mobility off road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.)

12.6 CONCLUSION

Developments in automotive engineering have always been used for terrain vehicles. It has to be considered that an off-road vehicle is an automobile with some applications for terrain mobility purposes. The engineers of terrain vehicles should apply general developments of present-day automotive engineering for the off-road vehicles of the future.

More recently, an increasing amount of work is now being done by computer simulation. Computer simulation enables engineers to evaluate the performance of equipment under virtual extreme conditions that could be dangerous, damaging, and expensive to perform on the test track. In a computer simulation, it is relatively easy to change the vehicle design parameters.

More than for road vehicles, computer modeling for off-road vehicles is essential.

The term modeling has two meanings for off-road vehicles:

- Modeling of the interaction between running gear and terrain
- Modeling of the vehicle (power train, suspension, running gear, etc.).

In the future, both models might be connected in combination with terrain maps. This model can show an off-road ride in a virtual world. The terrain ride can be seen from the position of an observer outside of the vehicle. Such models have been initiated by and will be improved further.

A next step could be a virtual test of the components and the whole vehicle during an off-road ride.

However, there is also a danger in virtual modeling. It must be considered that a virtual model is very close to reality but is not reality. Therefore, in the validation process, the virtual and the physical models have to be tested identically.

Electronic systems in road and off-road vehicles are becoming more and more common. *X-by-wire systems* are electronic systems without mechanical backup where the X stands for safety related applications, for example, steer by wire and brake by wire.

Hybrid vehicles [72,73] have a great chance as future propulsion systems for future off-road vehicles. Therefore, many attempts have been made to combine electric drives with conventional combustion engines.

The next generation of military ground combat vehicles may be equipped with hybrid electric drives to gain fuel efficiency, generate an abundance of onboard electric power, and reduce noise.

The use of an electric motor/generator in combination with an internal combustion engine offers improved fuel economy and reduced emissions as well as a simpler power train. Hybrid propulsion systems will be a serious alternative to conventional drives, especially for military terrain vehicles in wheeled or tracked concepts.

There is a principal difference between hybrid vehicles for wheeled and racked versions. For wheeled vehicles, two kinds of drive systems are possible:

- Electric motor installed in the hull
- Electric motor installed in the hub of each wheel [74]

The "in-hull installation" is more related to the conventional power train systems. The wheels of a mechanical power train system can be used. Disadvantages are the use of a cardan shaft and limited space for steering.

Conventional tracked armored vehicles have a very complicated mechanic–hydraulic steering system and could be substituted by electric motors. Another solution could be to use two electric motors directly connected for each track whereby very smooth steering is possible [75,76].

Many situations in the terrain cannot be presented in practical training. Therefore, it is necessary to simulate various situations. Nowadays, two-thirds of the driver training is done on a simulator. With this device, unusual and dangerous situations can be simulated. This saves time and cost.

Simulation must be as realistic as possible. Not only the image on the screen must look similar to the real terrain but also the movements of the simulator must follow the same motions as a real vehicle. Of course, also the driver cabin and all handlings devices must be identical with the original vehicle.

In the future, another philosophy, the modular concept for armored off-road vehicles, could take place. The chassis with hull power train and suspension will be produced separately from the body modules. The base vehicle operates independently from the modules. The modules are easily interchangeable and allow maximum flexibility for different mission requirements and multipurpose operations.

Another module system will be to build a low-weight basic tracked vehicle, which has a high maneuverability and can be air transported. For more dangerous situations, additional armor protection can be fitted easily. With this concept, different protection classes for different missions can be realized with one basic vehicle.

Due to the increasing cost constraints also for soft-skinned off-road vehicles, multifunctional equipments for armed forces and civil organizations are required. Different superstructures make the basic vehicle capable for a wide range of functions and missions.

On the other hand, the protection of the environment should not be forgotten. In the future, more emphasis should be put on the environment aspects. This is not only to reduce the exhaust gases but also to save the natural surface where terrain vehicles operate.

Military and civilian off-road vehicles are frequently driven in ecologically sensitive areas. Spilling of fuel and lubricants is sometimes unavoidable. In this respect, biodiesel could have the great advantage of low toxicity and high biodegradability.

Ground pressure exerted by tires or tracks can damage thin and not visible roots. Visible and larger roots suffer from the removal of their bark. The driver should not drive too near to trees and should avoid extreme acceleration and braking in such situations.

REFERENCES

1. Hohl, G.H.: Geländegängige Fahrzeuge, Script for the lecture at the Technical University Vienna, Vienna, Austria.
2. Bekker, M.G.: *Introduction to Terrain-Vehicle Systems*, University of Michigan Press, Ann Arbor, MI, 1969.
3. Yong, R.N.: *Vehicle Traction Mechanics*, Elsevier, New York.
4. Ford, R.: *Panzer von 1916 bis heute*, Karl Müller Verlag, Köln, Germany.
5. Bekker, M.G.: *Off-the-Road Locomotion*, University of Michigan Press, Ann Arbor, MI, 1960.
6. Bekker, M.G.: *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956.
7. Wong, J.Y.: *Theory of Ground Vehicles*, John Wiley & Sons, New York, 2001.
8. Wong, J.Y.: *Terramechanics and Off-Road Vehicles*, Elsevier, Amsterdam, the Netherlands.

9. Holm, C., Hefner, G.J., and Hintze, D.: The influence of shape and size of a penetration body on the pressure–sinkage relationship, *9th International ISTVS Conference*, Barcelona, Spain, 1987.
10. Karafiath, L.L. and Nowatzki, E.A.: *Soil Mechanics for Off-Road Vehicle Engineering*, Trans Tech Publications, Aedermannsdorf, Switzerland.
11. Hintze, D.: The influence of seasonal moisture changes on soil strength, *10th International ISTVS Conference*, Kobe, Japan, 1990.
12. Krick, G.: Wechselbeziehung zwischen starrem Rad, Luftreifen und nachgiebigem Boden, Thesis, Munich, Germany, 1971.
13. Shoop, S., Young, B., Alger, R., and Davis, J.: Winter traction using three different instrumented vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV, 1993.
14. Steiner, M.: Analyse, Synthese and Berechnungsmethoden der Triebkraft -Schlupf-Kurve von Luftreifen auf nachgiebigem Boden, Thesis, Munich, Germany, 1979.
15. Schmid, I.: Grundlagen der Geländegängigkeit—Wechselwirkung zwischen Rad und nachgiebigen Boden, Vorlesung des Institutes für Kraftfahrwesen und Kolbenmaschinen der Universität der Bundeswehr Hamburg, Hamburg, Germany, 1990.
16. Schmid, I.: Interaction of vehicle and terrain, *6th European ISTVS Conference*, Vienna, Austria, 1994.
17. Schmid, I.: Inflation pressure control to improve terrain trafficability: A model for simulation, *11th International ISTVS Conference*, Lake Tahoe, NV.
18. Hohl, G.: Correlation between slip-velocity-shear displacement for the prediction of drawbar pull-slip of flexible tyres, *12th International ISTVS Conference*, Bejing, People's Republic of China.
19. Holm, C.: Das Verhalten von Reifen beim Überfahren einer Spur auf nachgiebigen Boden und der Einfluss auf die Konzeption mehrachsiger Fahrzeuge, Thesis, VDI Fortschritsbericht 14/17, 1972.
20. Aubel, T.: Simulationsverfahren zur Untersuchung der Wechselwirkung zwischen Reifen und nachgiebiger Fahrbahn auf Basis der Finiten Elemente Methode; Thesis, University of German Armed Forces, Hamburg, Germany, 1994.
21. Aubel, T.: *The Interaction between the Rolling Tyre and the Soft Ground*.
22. Fervers, W.: FE-simulation of tyre-profile-effects on terrain mobility of vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.
23. Fervers, W.: Phänomene von Luftreifen und Geländeböden—Untersuchung mit FME. Universität der Bundeswehr—Institut für Kraftfahrwesen und Kolbenmaschinen, Thesis, 1999.
24. Takahashi, H.: Simulation of resistive forces acting on the bucket of wheel loader by using of DEM, *5th Asia-Pacific Regional ISTVS Conference*, Seoul, Korea.
25. Tanaka, H., Momotsu, M., Oida, A., and Yamazaki, M.: Simulation of soil deformation and resistance at bar penetration by the distinct element method, *12th International Conference of the International Society for Terrain-Vehicle Systems*, Beijing, People's Republic of China, China Machine Press.
26. Turnage, G.W.: Prediction of in-sand tire and wheeled vehicle drawbar performance, *8th International ISTVS Conference*, Cambridge, U.K., 1984.
27. Maclaurin E.B.: The use of mobility numbers to describe the in-field tractive performance of pneumatic tyres, *9th International ISTVS Conference*, Kobe, Japan.
28. Hohl, G.: A computer simulation model to predict speed v.s. time in off-road vehicles, *9th International ISTVS Conference*, Kobe, Japan.
29. Hohl, G.: Using a computer model to determine speed and acceleration of wheeled off-road vehicles, *5th European ISTVS Conference*, Budapest, Hungary.
30. Korlath, G.: A computer aided model to predict acceleration and maximum speed of tracked vehicles on unpaved ground, *13th International ISTVS Conference*, Munich, Germany, 1999.
31. VDI 2057 Standard May 1981.
32. ISO 2631-1 Standard 1978-01-15.
33. ISO 2631-1 Standard 1997-05-15.
34. ÖNORM ISO 2631-1; 2005-11-01.
35. Wolf, E.J.: *Using Absorbed Power Method to Evaluate Effectiveness of Selected Seat Cushions During Manual Wheel Chair Propulsion*, University of Pittsburgh, Pittsburgh, PA, 2003.
36. Lenzuni, P. and Lundström, R.: Guidelines for conducing laboratory measurements of the power absorbed by hand and arm; Vibration Injury Network; Appendix H3D to final Report May 2001.
37. Klanner, R.: Cross-country trucks for fast off-highway operation, *6th International ISTVS Conference*, Vienna, Austria, 1978.
38. Hoepke, E. et al.: *Nutzfahrzeugtechnik ATZMTZ-Fachbuch*, Vieweg Verlag, Berlin, Germany.
39. Nedvidek, H.: VW Iltis–ein Geländefahrzeug, *6th International ISTVS Conference*, Vienna, Austria, 1978.

40. Ogorkiewcz, R.M.: *Tank Technology*, Jane's Information Group, London, U.K.
41. N.N.: *Diesel Power in Military Applications*. Reprint from the MTU Supplement in *Military Technology*. Issue 4/86.
42. Krapke, P.W.: *Leopard 2*; Verlag E. S. Mittler Sohn, Bonn, Germany.
43. Aitzetmüller, H.: Steyr S-matic—The future continuously variable transmission for all terrain vehicles, *13th International ISTVS Conference*, Munich, Germany, 1999.
44. Pohlenz, J.: The new continuously variable ZF Eccom tractor transmission, *13th International ISTVS Conference*, Munich, Germany, 1999.
45. N.N.: Principles of Automotive Vehicles, Department of the U.S. Army Technical Manual, TM9-8000; TO36A-1-76, U.S. Army Material Command, Departments of the Army and the Airforce, Huntsville, AL.
46. Stockmar, J.: *Das große Buch der Allradtechnik*, Motorbuchverlag, Stuttgart, Germany.
47. Stelzeneder, F.X.: ADM a new drive train management, *6th International ISTVS Conference*, Vienna, Austria, 1978.
48. Hohl, G.: ADM—The automatic drive train system—A step forward in off-road mobility, *ISTVS Symposium*, Wroclaw, Poland, 1996.
49. Hohl, G.: The automatic drive train management system (ADM): Description and results, *7th International ISTVS Conference*, Ferrara, Italy.
50. Skoff, G.: Auswirkungen eines vollautomatisch gesteuerten Sperrmanagents auf den Antriebsstrang am Beispiel eines hochmobilen 6 × 6-Geländefahrzeuges. *VDI-Fortschritt Berichte, Reihe 12: Verkehrstechnik/Fahrzeugtechnik Nr. 351*.
51. Brügel, E.: Die Viscokupplung als Bauelement im Antriebsstrang, *ATZ*, 1986/12.
52. Stockmar, J.: Self-generated torques in all-wheel drive power train, *6th European ISTVS Conference*, Vienna, Austria, 1994.
53. Lechner, G. and Naunheimer, H.: *Automotive Transmissions*, Springer Verlag, Berlin, Germany, *Anti-Wheel-Spin Systems for Road Vehicles*, London U.K., 1988.
54. Stumpf, H.W. and Hohl, G.: Military off-the road tires, *FISITA 2000 World Congress*, Seoul, Korea, 2000.
55. Schlechter, H.P.: Design and performance of central tire inflation systems, *6th European ISTVS Conference*, Vienna, Austria, 1994.
56. http://auto.howstuffworks.com/self-inflating-tire2.htm
57. http://en.wikipedia.org/wiki/Run_flat_tire
58. Engeler, A.: Reifen mit Notlaufeigenschaften, Diploma Thesis, Technical University Vienna, Vienna, Austria, 1986.
59. Hohl, G. and Pächter, H.: Versuche an Reifen mit Notlaufeigenschaften, *FISITA Congress 1986*, Belgrade, Yugoslavia, 1986.
60. Hohl, G.: Tests of off-road tyres with emergency capabilities, *4th European ISTVS Conference*, Wageningen, the Netherlands, 1989.
61. FINABEL Recommendation 20 A 5, 1978.
62. Ageikin, Ia.S.: *Off-the Road Wheeled and Combined Traction Devices*, Oxonion Press PVT. Ltd., New Delhi, India.
63. Klanner, R.: Die Radaufhängung des geländegängigen Kraftfahrzeuges, Thesis, Technical University Vienna, Vienna, Austria, 1943.
64. Hohl, G.H.: The influence of the suspension system on the mobility of wheeled off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993.
65. Hohl, G.H.: An overview of suspension systems of wheeled off-road vehicles, *10th ISTVS Conference*, Vienna, Austria, 1994.
66. Junker, H. and Seewald, A.: A theoretical investigation of an active suspension system for wheeled tractors, *8th International ISTVS Conference*, Cambridge, U.K., 1984.
67. Deakin, A.J., Crolla, A., and Roberts, D.S.A.: Development of semi-active suspension for combat support vehicles, *FISITA 2000 Congress*, Seoul, South Korea, 2000.
68. Nell, S.: Experimental solution of an unsophisticated two state semi-active damper, *Journal of Terramechanics*, 31(4), 227–238.
69. Crolla, A., Deakin, A.J., and Roberts, D.S.A.: Rapid control prototyping for semi-active suspension system to be fitted to heavy off-road vehicles, *European Automotive Congress*, Bratislava, Slovak Republic, 2001.
70. Beno, J.H. and Weeks, D.A.: Active suspension technology development for high performance off-road vehicles, *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.

71. Crolla, D.A.: A comparison of axle and independent suspension for off-road vehicles, *11th ISTVS Conference*, Lake Tahoe, NV, 1993.
72. Rosenfeld, R.L.: Lessions from electric vehicle technology development, *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.
73. Sorg, J.: Electric drives a viable alternative for automotive drives? *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.
74. Goldsack, S. et al.: Hub mounted electric drive project at Qinetio and MST (U.K.), *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.
75. Thomson, R.W.: Electric drives for tracked vehicles, *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.
76. Walter, A.: Safety requirements for the electro-mechanical-transmission (EMT), *4th All Electric Combat Vehicle Conference*, Noordwijkerhout, the Netherlands, 2002.
77. Ellis, C. and Bishop, D.: *Military Transport of World War I*. Brandford Press Ltd., England, U.K.
78. Hohl, G.H. and Corrieri, A.: Basic considerations for the concepts of wheeled off-road vehicles. *FISITA World Automotive Congress*, Seoul, Korea, 2000.
79. Ruff, K. and Jakobs, R.: Interactive simulation of vehicle movement in terrain, *6th European ISTVS Conference*, Vienna, Austria, 1994.
80. Schwanghart, H. and Zweier, A.: Influence of tread on drawbar pull of tires for small off-road vehicles, *11th International ISTVS Conference*, Lake Tahoe, NV.
81. Maclaurin, E.B.: A rig for testing the soft soil performance of tracked systems, *8th International ISTVS Conference*, Cambridge, U.K., 1984.
82. Kutzbach, H.D.: Investigation on tractor tyres test stand and results, *6th European ISTVS Conference*, Vienna, Austria, 1994.
83. Hohl, G.: A method to determine coefficient of traction-slip relationship of tyres in "in-situ" tests, *9th International ISTVS Conference*, Barcelona, Spain.
84. N.N.: *Engineering Design Handbook*, Automotive Series, Automotive Suspensions, AMCP 706–356, Headquarters, U.S. Army Materiel Command, Huntsville, AL.
85. Hohl, G.H. and Peschak, B.: A model to predict the overcoming of geometrically defined obstacles by a two-axle wheeled vehicle, *13th International ISTVS Conference*, TU München, Munich, Germany, 1999.
86. Antonov, A.S.: Militärfahrzeuge Band 2. Militärverlag der Deutschen Demokratischen Republik, Berlin, Germany.
87. Ageikin, Ia.S.: *Off-the Road Mobility of Automobiles*, Oxonion Press PVT. Ltd., New Delhi, India.
88. Hohl, G.: Ride comfort of off-road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.
89. Merhof, W. and Hackbart, E.M.: *Fahrmechanik der Kettenfahrzeuge*, Leuchtturm, Germany.
90. NN.: Brochure of ZF Company, Friedrichshafen, Germany.
91. Braess, H.-H.: *Handbuch Kraftfahrzeugtechnik*, Vieweg Verlag, Berlin, Germany.
92. http://www.actiontireco.com/Scripts/tiredefinitions.asp
93. Folder Diehl Company, Modesto, CA.
94. Hohl, G. and Heimberg, N.: The duro 4 × 4 and 6 × 6—A new concept for a family of soft-skinned and protected off-road vehicles, *EAEC 2001 Congress*, Bratislava, Slovak Republic.
95. Folder of Timony Company.
96. Folder of Steyr Spezialfahrzeuge Company.
97. Spielberger, W.J.: *Von der Zugmschine zum Leopard 2*, Bernhard & Graefe Verlag, München, Germany.
98. Holman, T.J.: Rotary hydraulic suspension dampers for high mobility off road vehicles, *8th International ISTVS Conference*, Cambridge, U.K., 1984.
99. Bollmann, St.: Transmission system of tracked vehicles, *6th European ISTVS Conference*, Vienna, Austria, 1994.
100. Heisler, H.: *Vehicle and Engine Technology*, Traction Properties of Off Road Vehicles, SAE International, Warrendale, PA.
101. Reimpell, J.: *Fahrwerktechnik*, Vogel Verlag, Würzburg, Germany.
102. Hohl, G.: Torsion bar spring and damping systems of tracked vehicles, *Journal of Terramechanics*, 22(4), 195–203, 1986.
103. Schmid, I.: Kettenfahrzeuge. Vorlesung des Institutes für Kraftfahrwesen und Kolbenmaschinen der Universität der Bundeswehr Hamburg, Hamburg, Germany, 1990.

13 Motorcycles and Three-Wheeled Vehicles

Robin S. Sharp

CONTENTS

13.1 Introduction 477
13.2 Motorcycles 477
13.2.1 Frame 478
13.2.2 Rear Suspension 479
13.2.3 Front Suspension 483
13.3 Three-Wheeled Vehicles 487
13.3.1 Nontilting Three-Wheelers 487
13.3.2 Tilting Three-Wheelers 487
13.3.2.1 Tilt Control and General Handling Properties 488
13.3.2.2 Stability and Response Properties of the Ariel 3 490
13.4 Conclusions 493
References 494

13.1 INTRODUCTION

Basic architectural features of motorcycles will be discussed first, followed by considerations of rigid-framed and tilting three-wheelers. Motorcycle architectures derive from those of bicycles, while the rigid-framed three-wheeler is essentially a car in engineering terms. Tilting three-wheelers seem to have more in common with the motorcycle than with the car, but there are elements of both involved. Tilting may be rider controlled or actuator controlled through motion feedback. In the latter case, tilting moments applied by actuators must be reacted on a nontilting part of the vehicle. The influence of these moments on the motions of the nontilting part may be important and should be considered in any design study.

13.2 MOTORCYCLES

Essential to a conventional motorcycle are two wheels nominally in a single plane, with a steering system allowing the front wheel to steer freely around an inclined axis. The longitudinal separation of the wheels defines the wheelbase. The inclined steering axis is structurally convenient, allowing the steering axis to lead, at ground level, the front tire/ground contact center and the handlebars to be conveniently located for the rider. The inclined steering axis is also an important determinant of the stability, control, and response properties of a motorcycle. At low speeds, it is evident that an uncontrolled motorcycle will be unstable in roll, but it transpires that it will typically self-stabilize at a modest speed, as described in Chapter 35. The freedom of the steering system is important to a motorcycle's autostabilization, since this autostabilization depends on the front system steering into a fall, thereby introducing tire forces, through cambering and sideslipping of the front tire, which exert a rolling moment around the machine's mass center necessary to counter the original fall [1]. Those aspects of the design of the motorcycle that influence the self-steering behavior deserve

particular attention. Such aspects include the steering rake, the mechanical trail, and the front wheel gyroscopic torques. Steering the rear wheel appears superficially to be feasible, but dynamic analysis and experimentation [2,3] show it not to be practical. Steering both front and rear wheels has been researched [4], but it does not appear to be advantageous.

The possession of a free steering system leads to the possibility of oscillatory instabilities. When the self-steering is too vigorous, overshoot and oscillations result. Under some running conditions, it may be essential to restrain the steering freedom, to some extent, by means of a damper or other device. Compromise between what is best for one set of conditions and what is best for another may well be necessary. The main oscillatory modes are weave and wobble modes, both of which can be observed near to straight running at constant speed (see Chapter 35). Variants of these modes occur in cornering and in braking or driving, especially under extreme circumstances typical of racing. Oscillatory modes with higher frequencies and of less certain character are often called by the names patter, chatter, or judder [4,5]. An example demonstrated in simulation can be found in Ref. [6], but it is not known how close to real events these results come.

A main structure to connect the wheels together, an engine and transmission system, suspensions, brakes, fuelling, lights, etc., and controls are needed. The main mass of the motorcycle is the power train, an assembly of engine, clutch, and gearbox, which is substantially rigid and which invariably is placed between the wheels, with the rider on, as opposed to in, the motorcycle. The wheelbase is normally about as short as is convenient from a packaging point of view, which implies that the mass center is quite high and front-to-rear load transfers in accelerating or braking are severe. With high-performance machines, in fact, it is not uncommon for longitudinal maneuvering to be limited by wheelie or stoppie, as can be easily observed at the race track. Longer wheelbases, lower riders, and more than usual weather protection have been tried. The Quasar, for example, had a 1.96 m wheelbase, compared with norms for tourers of 1.47–1.52 m and for racing machines of 1.27–1.4 m. Many details can be found in Ref. [4].

The compliance of the structure, the steering geometry, any steering restraints, gyroscopic influences of spinning wheels and power train parts, suspension design details, speed range, performance priorities, tire characteristics, aerodynamic properties, and rider dynamics all interact with each other to determine the performance of a motorcycle/rider combination. The extent to which the parts can be separated from each other and discussed individually is limited by this strong interdependence. Such discussion must be directed mainly at understanding the rather general influences of design features on behavior.

13.2.1 Frame

The main necessary property of the frame, in terms of performance of the motorcycle, is torsional stiffness. If the frame is too compliant, the weave mode especially will become underdamped or even unstable as speed increases, and the motorcycle will become dangerous. The stiffness required depends strongly on the speed range of the machine. High-performance machines need to be stiffer than others. As the frame stiffness increases, the damping of the weave mode at high speed improves, but the speed at which the wobble mode is minimally damped increases also. There comes a point at which the wobble mode is a greater problem at high speed than the weave mode is. The wobble mode stability is increased very effectively by a steering damper, but this has a bad influence on the weave (see Chapter 35). Racing machines consequently can be characterized as having stiff frames and needing steering dampers. If they wobble, it is at high speed. Touring machines are generally more compliant and have less need for a steering damper, and, if they wobble perceptibly (commonly referred to as "shaking their heads"), it is at more moderate speeds. Anecdotal accounts suggest that such modest-speed wobbling is much less vigorous than the high-speed variety. Usually, it can be controlled by the rider. The idea of planned compliance can be found in Ref. [7].

Presuming that the power train is relatively heavy and that it is located between the wheels, in simple terms, the frame structure can be considered to be in two parts. One part joins the power

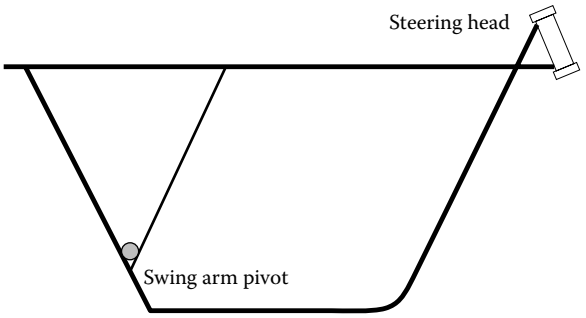


FIGURE 13.1 Diagrammatic representation of a perimeter frame. Such frames are normally made of small-diameter steel tubes and gussets, and they can be constructed with very little manufacturing infrastructure. Most motorcycles coming after the widespread adoption of swinging arm rear suspensions have had such frames. Many examples can be found in. (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com); Wilson, H., *The Encyclopedia of the Motorcycle*, Dorling Kindersley, London, U.K., 1995.)

train mass to the front wheel, while the other joins it to the rear wheel. Usually, the structural route to the front wheel contact patch entails joining the engine to the steering head and the steering head to the contact patch, through the front suspension mechanism, the wheel bearings, and the wheel and tire compliances. Normally, also, the structural route to the rear wheel contact patch entails joining the power train to the swing arm pivot and the swing arm pivot to the contact patch through the swing arm, wheel bearings, and wheel and tire compliance. The point of thinking about the frame structure in this way is that the main sources of external force on the motorcycle are the contact patches and the main source of inertial resistance to motion in response to those forces is provided by the power train. Further, the power train itself may be used as part of the structure, in which case the main structure literally is in two parts, joining engine to steering head at the front and gearbox to swing arm pivot at the rear. The compliances of the structural parts and of the tire can be thought of as in series with the frame compliances, so that there is little point in stiffening the structural parts further, once they are much stiffer than the tire carcass.

It is relatively easy to achieve the necessary torsional stiffness in the rear one of these substructures, since no steering is involved and the load paths are quite direct. Calculations have shown that diminishing returns for increasing torsional stiffness of a swing arm set in at about 12 k Nm/rad [8], while the stiffness of a contemporary superbike swing arm is around 50 k Nm/rad [5]. Provided that the design details are good [4], it appears not difficult to obtain high quality of performance in this area. The front structure is a different matter. The load path is usually long and tortuous, and both steering and suspension joints are included in the path. With conventional geometry, it is difficult to make the structural parts as stiff laterally as the front tire carcass. As a consequence, many variations on the conventional steering arrangements have been used [4,5], with strong implications for the frame design.

Some alternative frame structural concepts are illustrated in Figures 13.1 through 13.5, the first four of which rely on joining the steering head to the swing arm pivot while supporting the engine, rider, etc. The remaining one employs the power train as part of the structure and may include only subframes for the front and rear. If the mechanical balancing of the engine is poor, using the engine as part of the structure may well lead to vibration problems.

13.2.2 Rear Suspension

The swinging arm rear suspension is now almost universal. It is considered not worthwhile to discuss other possibilities. Torque steer is an issue with swing arm suspension. Ideally, neither driving nor braking will lead to any steer deflection of the rear wheel. In the braking case, ensuring

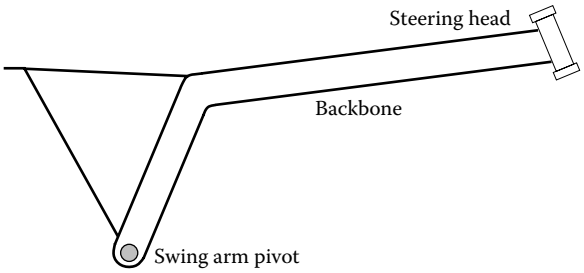


FIGURE 13.2 Diagrammatic representation of a backbone frame. Torsional stiffness comes mainly from one large-diameter tube, which is bent to clear the power train as necessary. (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com).)

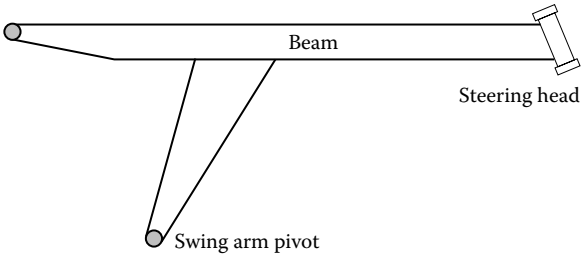


FIGURE 13.3 Diagrammatic representation of a beam frame. Such a frame is normally made of steel pressings, implying a strong connection with large-scale production. The Ariel Leader and Arrow are examples. (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com); Wilson, H., *The Encyclopedia of the Motorcycle*, Dorling Kindersley, London, U.K., 1995.)

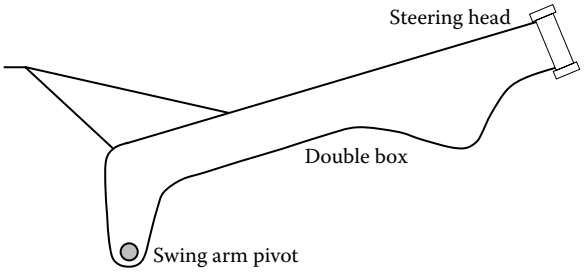


FIGURE 13.4 Diagrammatic representation of a double-box frame. A pair of closed section boxes straddle the top of the power train. These modern frames are normally made of aluminum alloy. (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com); Cameron, K., *Sportbike Performance Handbook*, MBI Publishing, Osceola WI, 1998; Wilson, H., *The Encyclopedia of the Motorcycle*, Dorling Kindersley, London, U.K., 1995.)

lateral symmetry is one approach to eliminating torque steer, implying the need for brake discs and calipers on both left and right. For driving by shaft, left/right symmetry would be perfect but is unlikely to occur, due to the drive arrangements. For driving by chain, the ideal would be to have the lateral elastic center of the swing arm a little inside the line of the chain, so that the swing arm simply translates longitudinally, without yawing, when the chain tension changes and the corresponding, much smaller, longitudinal tire force is generated. These ideals may be mutually exclusive or impractical, so that a small amount of torque steer may have to be allowed. Obviously, high stiffness will help to lower the influence of any geometric imperfections in a design.

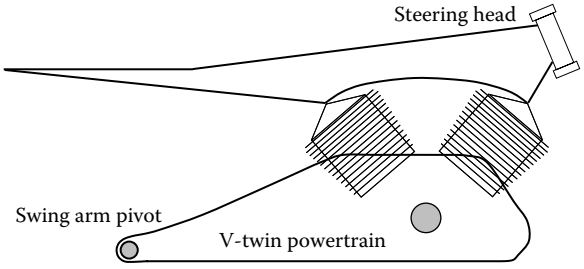


FIGURE 13.5 Diagrammatic representation of a two-part frame, utilizing the power train as the main structural component. Such a frame concept may involve small-diameter tubing with triangulation to make a "space frame." (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com); Wilson, H., *The Encyclopedia of the Motorcycle*, Dorling Kindersley, London, U.K., 1995.)

Most modern motorcycles have a single spring/damper unit at the rear, a so-called monoshock, with a multilink system connecting it to the swing arm and the frame. The suspension unit is likely to include elastic bump and rebound stops, which give the system a "hardening" property (the spring stiffness increases as the deflection from nominal increases) as the spring deflection increases toward the travel limit. Much is often made of the possibility of using the linkage design, or the spring design itself, to shape the suspension stiffness [7], but it seems likely that such shaping will be overwhelmed by the limit stops and therefore is not important. Also, analysis often reveals a substantially linear characteristic to the mechanism involved [10].

Monoshock arrangements are many and varied. Examples can be found in Refs. [4,5], and several different geometries are analyzed in Ref. [11]. The main considerations in selecting a monoshock design appear to be minimizing part count, cost, and mass; packaging the components neatly in the space available; and reducing the need for main structure aft of the swing arm pivot line.

Under driving or braking, the attitude of a motorcycle will change in response to the tire longitudinal forces responsible for the motion. The forces act at ground level, and the inertial reactions to the forces are all above ground level, so that, corresponding to an acceleration, there must be a load transfer from front to rear tires and conversely (Figure 13.6).

Rear suspension and transmission system geometries influence the attitude changes and the extent to which the limited suspension travel is used to accommodate accelerating or braking. Let us first consider a swing arm geometry, which involves a more or less horizontal swing arm at the nominal state, combined with a single-jointed shaft drive. The shaft joint needs to be aligned with

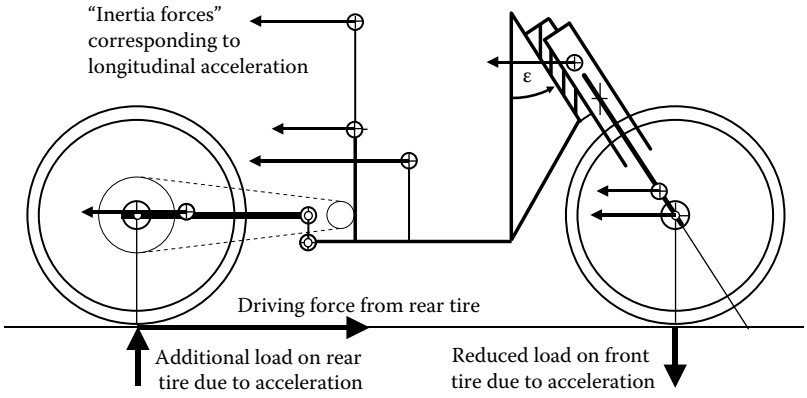


FIGURE 13.6 Illustration of load transfer from front to rear under acceleration caused by drive thrust at the rear tire contact patch. The load transfer must occur to account for the equilibrium in pitch of the whole system, irrespective of the internal suspension and transmission details.

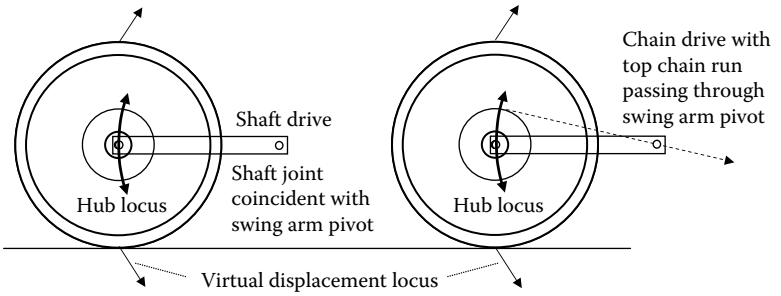


FIGURE 13.7 Shaft and chain drive rear suspensions with strong antisquat property. In each case, the swing arm is horizontal. A single-joint shaft-driven wheel and a chain-driven wheel with top chain run passing through the swing arm pivot are equivalent systems with respect to squatting, for small perturbations.

the swing arm pivot. Since we are dealing with a multibody system, we must use the principle of virtual work or virtual power to examine equilibrium states [12]. Imagine a virtual displacement in which the swing arm rotates toward rebound (positive), while the transmission system does not rotate in spin at all, implying that no work is done by the drive system. The drive then implies that the road wheel rotates with the swing arm as if the road wheel and the arm were one body. In the virtual displacement, the wheel center moves vertically, while the tire contact center moves in a direction perpendicular to the line joining it to the swing arm pivot. The longitudinal driving force moves along its line of action, and this force pulls the suspension into rebound, opposing the motion in bump corresponding to the load transfer. The system has a strong antisquat character, the more so the shorter is the swing arm [13]. A motorcycle with this layout may lift perceptibly when accelerating hard, the rear rising due to the suspension and transmission geometry, the front rising due to the load transfer from front to rear [4]. If we replace the shaft drive with chain drive and we consider the special case in which the upper chain line passes through the swing arm pivot, the argument and the conclusion are the same as before. The two cases are illustrated in Figure 13.7.

If an additional joint is included in the drive shaft and the swing arm is replaced by a parallelogram linkage as shown in Figure 13.8, the locus of the tire contact center in a virtual displacement in which the drive shaft does no work is vertical, since the road wheel simply translates without rotating. Further, if the driving sprocket of the swing arm, chain drive system is raised from its former position so that the upper chain run is parallel to the swing arm and horizontal, the virtual displacement in which the drive system does no work similarly involves the road wheel in translating without spinning in an absolute sense, when the swing arm moves in rebound. The road wheel has a rotation relative to the swing arm, which is equal and opposite to that of the swing arm. The contact center moves vertically, and the driving force does no work on the system. In each of these cases, the driving force exerts no kinematic influence on the suspension, which has a zero antisquat property.

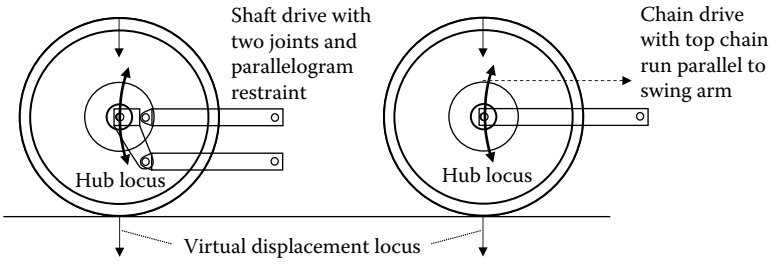


FIGURE 13.8 Shaft and chain drive rear suspensions with zero antisquat property. Generalizing the results, as the declination of the top chain run to the front increases, so the virtual displacement loci change and the degree of antisquat increases.

Thus, with the most common swing arm suspension with chain drive, the system has an antisquat character, which becomes stronger as the upper chain run declines to the front more steeply. These arguments are not changed fundamentally by inclination of the swing arm.

Rather intricate measurements or simulation results are necessary to illustrate antisquat properties more fully [11]. The study of the transient behavior, in particular, is likely to be ineffectual unless sophisticated methods are used. The results of [11] confirm the earlier findings qualitatively and add to them quantitatively. Those results also show that there is a price to pay for using antisquat kinematics to control attitude changes in acceleration, which is that transient load transfers are made more rapid and oscillatory than they otherwise would be (see Ref. [13] in addition). The rider's task of exercising good control while carrying out a serious maneuver is made more difficult by such rapidly changing tire loading.

The corresponding antidive problem in braking is qualitatively similar but is not such a big issue quantitatively since braking the rear wheel of a motorcycle usually needs to be done with reserve. The short wheelbase and the high-mass center imply that the load will come off the rear wheel almost entirely in a serious deceleration and, consequently, very little braking force can be generated there. The antidive kinematics depend on the caliper mounting arrangements, as previously the properties depended on the drive arrangements. Mostly, calipers are anchored on the swing arm, yielding the same character in braking as depicted for driving in the left-hand case of Figure 13.7. With such geometry, rear braking only will give a pitch down response by virtue of the front suspension compressing corresponding to the load transfer, while the swing arm's suspension movement is very small.

A notable feature of the rear suspension is the extent of the load change that typically occurs, with variations in operating circumstances. First, the difference between a static load of one small person and of two large people and some luggage is large in percentage terms. Secondly, the bank-to-turn cornering behavior implies that the suspension loading in a near limit turn is about 1.4 times the static load. The simplest device to cope with this large range of variation is the common manual adjustment of spring preload. A better but more expensive alternative is self-leveling, usually by means of a gas spring, a pump, and valves, for example, the Boge Nivomat unit [4]. A third way worth mentioning is the manually adjustable mechanical advantage or variable geometry system used by Velocette [4]. Such a variable geometry system can, in principle, be automated to give an active, low-power suspension with particular advantage in the application due to the special problems described. Such a system can also be designed to add damping usefully to the potentially problematic weave mode of a high-performance machine [14].

13.2.3 Front Suspension

The numerically dominant front suspension type for motorcycles is the telescopic fork. Typically, the fork legs are held in a "triple clamp" at the steering head with a small offset between the steering axis and the fork axis. The steer axis and the fork axis are normally parallel. Springs and dampers are contained within the forks and constitute part of the steering assembly. The sliding joints in the forks are quite critical components. They are required to provide friction-free movement, to accommodate bending loads especially in braking and to seal against dirt ingress and damper fluid leakage. The steering inertia is quite low, since the masses are concentrated near to the steering axis, but the structure is far from ideal, since the lateral support at the steering head is so far removed from the contact center. Stiffening of the fork structures conflicts, to some extent, with free movement at the joints and exaggerates the need for precise manufacture or fitting of the parts. Such forks have negative antidive kinematics when the brake calipers are anchored to the lower forks, depending on the fork rake. This is because the contact center moves rearward corresponding to a suspension displacement in bump, in which the braking torque does no work. Most of the available suspension motion may be used up in heavy braking. Increasing fork compression damping in response to front braking torque has been used as a means of reducing dive [7], but the practice is not common on current machines.

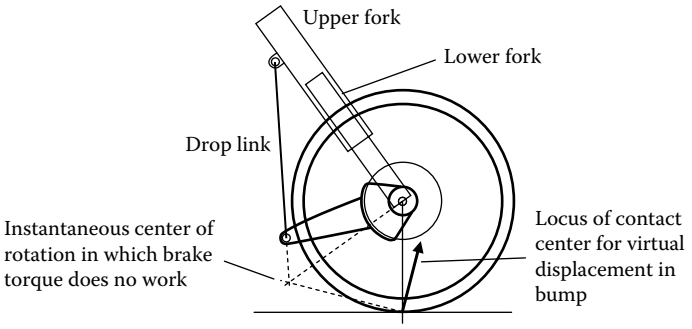


FIGURE 13.9 Telescopic fork with brake caliper rotation controlled by linkage to upper fork. The negative antidive property of the standard system, with caliper attached to fork lower leg, is replaced by antidive here. (From Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003. http://www.tonyfoale.com)

If the rotation of the brake calipers is controlled by linking to the upper forks, antidive can be designed into a telescopic fork (Figure 13.9), but it will have the same implications, as mentioned earlier, that front braking will lead to a more rapid and oscillatory change of loading of the front tire than with no antidive, challenging the rider's control capabilities to obtain vigorous deceleration smoothly. In an actual braking event, the geometry will change as the maneuver progresses, such that simulation is needed to get a full picture of the transient and the final equilibrium state [11,13]. Torque steer under braking is an issue as before. The solution by symmetry implies the need for similar discs and calipers on left and right forks.

Leading or trailing link systems are pin-jointed alternatives to the sliding joints of the telescopic fork, having potential kinematic advantage in terms of reduced friction and elimination of bending loads [4,5,9], but the stiffness-to-mass ratio properties are likely to be worse than with telescopics, and the distribution of mass away from the steer axis implied will increase the steering inertia, bringing down the wobble mode frequency and making weave and wobble conflicts greater than before and probably harder to resolve.

Continuing with systems in which the steering parts are exterior with respect to the wheel and tire, the point in the center plane just above the wheel is a key point structurally. Lateral support for the wheel assembly at this point is advantageous, since the distance from the support point to the contact center is minimized in comparison with any other points in the class. Various alternatives to the telescopic fork system use this key point. Typically, a ball joint that helps to define the steer axis will be located on a bridging member between the lower fork legs [4,7,9]. The joint can be associated with a McPherson strut-like Telelever (Figure 13.10) or a double-wishbone-like Hossack or Britten system [4,7] (Figure 13.11). The Telelever system retains the sliding joints of the telescopic forks, but the bending moments from braking are reduced. The antidive kinematics are quite different from those of the original. Structurally and kinematically, the Hossack design appears advantageous, although the steering linkage to the handlebars is more elaborate than normal. Both systems are good from the point of view of steering inertia, with one spring/damper unit not steered.

If we now consider front suspension and steering systems in which the wheel interior space is employed for parts, the essential feature is the location of a steering joint in the central plane and near to the wheel center. There are two ways in which this can be done. One of the ways involves deeply dishing the wheel, in car style, to make room for a wheel carrier or upright to have a spherical joint at the required location. There must also be a second joint on the upright in the wheel center plane to define the steering axis, and this second joint may be placed inside the wheel's interior space or above the wheel's orbit (Figure 13.12). The systems are sketched from the side in Figure 13.13.

The alternative way to locate a joint near the wheel center is to mount the wheel on a two-sided frame with a hollow cylinder at its heart, using large-diameter wheel bearings running on the

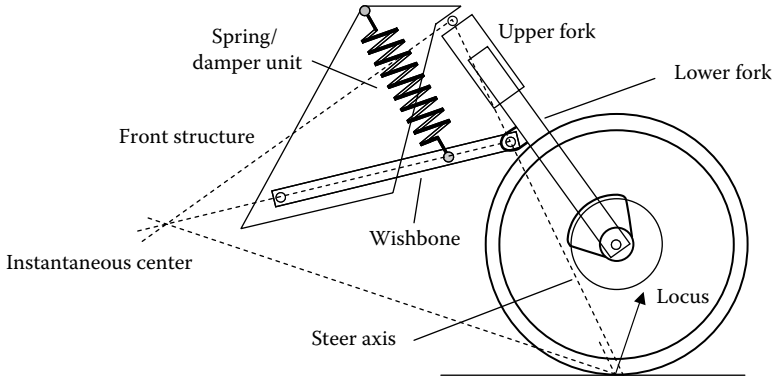


FIGURE 13.10 Telelever front suspension and steering system with ball-jointed support for the wheel assembly just above the wheel.

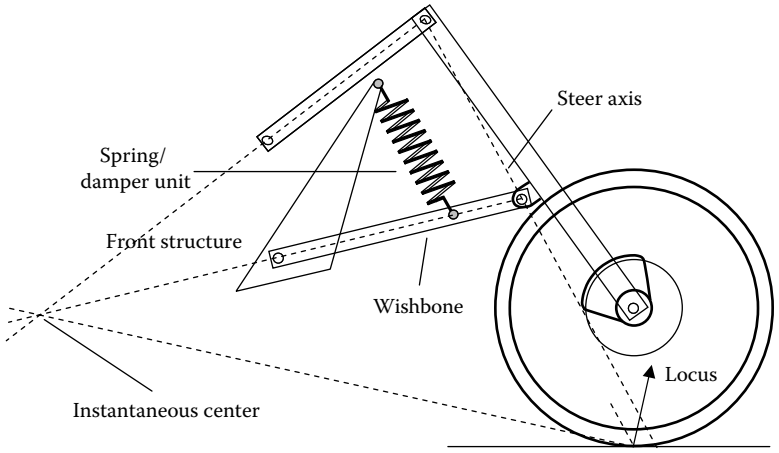


FIGURE 13.11 Hossack-type front suspension and steering system with ball-jointed support for the wheel assembly just above the wheel.

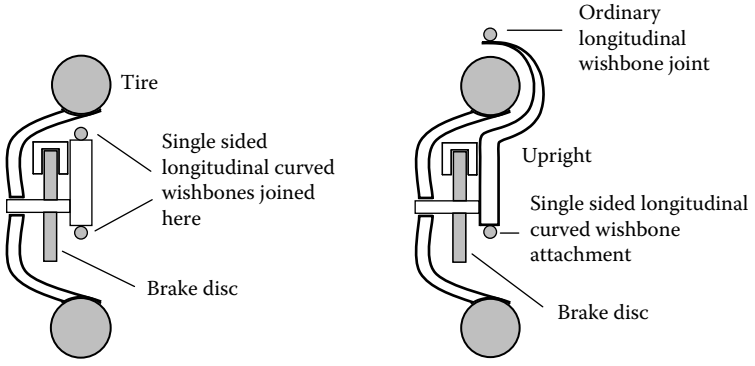


FIGURE 13.12 Diagrammatic illustration of front suspension arrangements using the wheel interior space for structural elements. On the left, both joints are interior, and both longitudinal suspension arms need to be curved to clear the tire on full lock. On the right, the upper suspension arm is exterior and is relatively unrestricted.

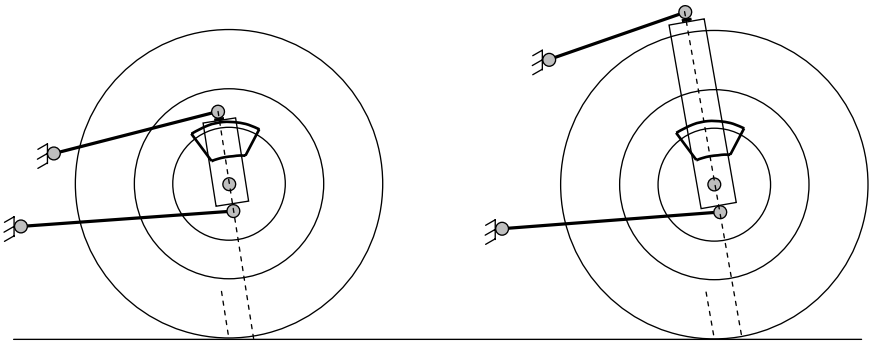


FIGURE 13.13 Diagrammatic side view of suspension systems shown in Figure 13.12, illustrating the freedoms to set the rake angle, the trail, and the antidive properties as desired. Note the implications of the choices made for the main frame.

hollow drum. The interior of the drum is used to locate the required joint. The lower suspension arm that relates to this joint may be two sided and must support a beam that passes through the drum. The drum must have sufficient diameter relative to its width to allow a reasonable steering lock (Figure 13.14). Note that the Bimota Tesi design dispenses with the upper wishbone and joint by using a revolute or pin joint at the hub interior and reacting the brake caliper force through a separate tension link [4,9]. Substantially, the same design freedoms as illustrated in Figure 13.13 exist. Several photographic illustrations of real systems of the types described are included in Refs. [4,9].

These so-called hub-center-steered systems shorten the load path from the front tire contact patch to the power train and thereby offer structural advantages. However, they all imply a restricted steering lock, due mainly to the need for the lower suspension member to clear the wheel and tire when the steering is on full lock and to not be so wide at the base as to ground in vigorous cornering. The restricted steering lock is more likely to be acceptable in racing than in a road machine.

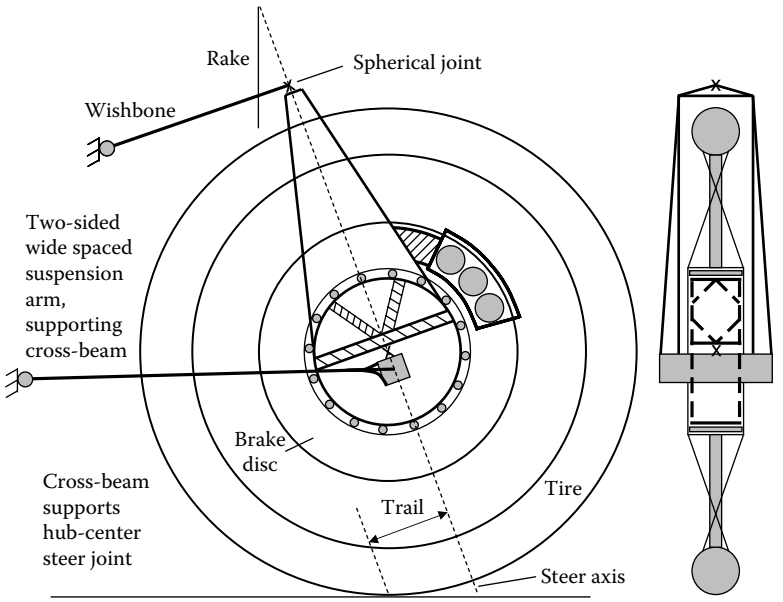


FIGURE 13.14 Location of steering joint near to wheel center inside a hollow cylinder supporting large-diameter wheel bearings. Beam and cylinder, and lower suspension arms and tire, must be shaped to allow a reasonable steering lock.

Once a motorcycle is traveling at reasonable speed, the steer angles used for control are usually small, so only the very low-speed behavior is prejudiced by limited lock.

13.3 THREE-WHEELED VEHICLES

Three-wheeled vehicles will be divided into rigid-framed or nontilting and pivot-framed or tilting classes. Essentially, the former class has more in common with the car than the motorcycle, and only a brief account will be given. Tilting three-wheelers include machines that are free tilting and steer much like a motorcycle and those with steering responses modified by an active tilt-control system. For ease of reference, we will use the designations 1F2R to represent a vehicle with one wheel at the front and two at the rear and 2F1R for the obvious alternative.

13.3.1 Nontilting Three-Wheelers

It is substantially true of both 1F2R and 2F1R designs that the total load needs to be shared roughly equally between the wheels statically. This requirement stems from the need to maneuver both longitudinally and laterally to a reasonable level while keeping the tires in contact with the ground. If all three tires are similar, 1F2R versions will tend to limit oversteer and static instability in yaw, while 2F1R versions will have the opposite tendency. This kind of behavior stems from the fact that all the lateral load transfer in cornering must occur at the pair of wheels, implying that the limit turning of a 1F2R on a good surface involves about one-third of the vehicle weight being supported by the front tire and about two-thirds being supported by the outside rear tire, with hardly any load on the inside rear tire. The rear axle is likely to saturate first as the cornering effort increases, despite that the front tire will camber with the vehicle body roll. A converse argument applies to the 2F1R design. This likely yaw imbalance of the three-wheeler is a fundamental disadvantage of the type.

Another disadvantage is that the overturning potential is significantly greater than with a corresponding four-wheeler, due to the near-zero track width at one end of the vehicle. To a certain extent, the nontilting three-wheeler has a greater need for width of track at its two-wheeled end, compared with a conventional car, to contain the rollover problem. Locating the mass center nearer to the two-wheeled end will reduce the overturning problem but will exacerbate the yaw imbalance problem. These disadvantages are not much offset by reduced weight or cost. Rigid-framed three-wheelers that have had a commercial life have generally owed it to legislation, giving them tax advantages in comparison with four-wheelers. A low-mass center reduces the severity of the problems mentioned, but that is not technically novel or interesting, since it is true of two-track vehicles generally.

13.3.2 Tilting Three-Wheelers

One possible purpose for a three-wheeler is that it behaves much like a two-wheeler but has a much increased potential for including weather protection, crashworthiness, and luggage carrying. In such a case, the rolling motion of the vehicle body is likely to be controlled by the rider, in the manner of the motorcycle. A potentially useful way of thinking about the roll control of a tilter is furnished by considering the vehicle to consist of an inverted pendulum pivoted longitudinally to a nonrolling trolley, illustrated in Figure 13.15. The Ariel 3 [15,16] was a commercial moped with this form of construction, as also were the Honda Gyro, the Honda Canopy, and the Honda Stream [16]. A more recent vehicle of this basic design is the GM Lean Machine [16]. The height of the joint between nonrolling and rolling sections is a design issue, as is the inclination of the joint to the horizontal, influencing roll steer. Locating the rolling joint near to ground level enhances the influence of tilting on the lateral load transfer, which is one of the main features. The Ariel 3 included a weak torsion bar in the pivot joint, so that the tilting frame would remain upright when the rider dismounted. Other designs have a lean locking brake, for operation by the rider at low speeds [16].

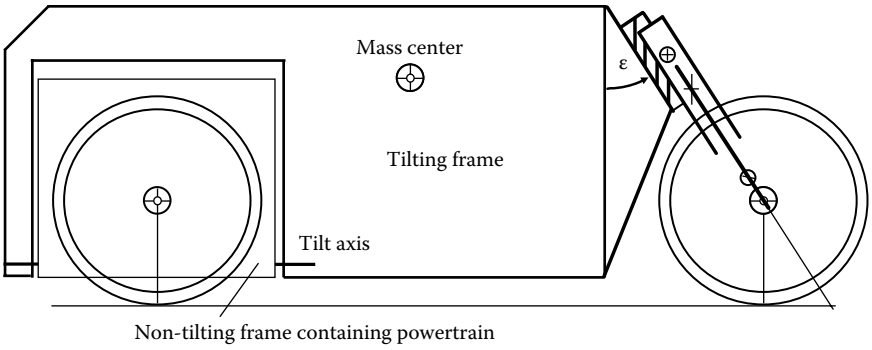


FIGURE 13.15 A diagrammatic view of a trolley and inverted pendulum type of tilting three-wheeled vehicle, in which the front is like a motorcycle and the rear more like a car. A rolling joint links the two frames together. It is shown horizontal here, but it may be inclined, leading to some roll steer.



FIGURE 13.16 Millennium Tracer all wheel tilting prototype (From http://www.members.iinet.net.au).

With this kind of vehicle design, the rear tires develop lateral forces primarily by sideslipping, while the front tires do so by cambering. There is a strong case for using car tires on the rear and a motorcycle tire on the front.

An alternative kinematic arrangement involves tilting all three wheels as the structure rolls. The Australian Millennium Tracer (Figure 13.16) and the U.S. Calleja are examples [16]. The Mercedes Life Jet experimental vehicle may also be included (Figure 13.17). These vehicles require a suspension design associated with the wheel pair, which provides bump stiffness, with very little roll stiffness. Linking the left and right wheels with a freely rolling transverse beam with its pivot axis restrained vertically by springing and damping will suffice, and putting a rider-operated or automatic brake on the pivot allows elimination of the body roll at parking speeds. In these cases, all the tires operate at high camber angles and need to be of motorcycle type. The Tracer and the Life Jet are 2F1R types, while the remainder are 1F2R.

13.3.2.1 Tilt Control and General Handling Properties

Free-tilting or largely free-tilting vehicles have their rolling motions controlled by the rider, and the straight-line stability and response properties are similar to those of motorcycles [15]. This implies that they have potential appeal only to people who already can ride a bicycle. The overturning problem that looms large with narrow two-track vehicles is dealt with effectively by the rider



FIGURE 13.17 Mercedes Life Jet all wheel tilting prototype leaning into a turn. (From Foale, T., *Some Technical Aspects of Tilting Trikes*, Tony Foale Designs, Alicante, Spain, 1999, pp. 24, http://www.motochassis.com with kind permission of Tony Foale.)

knowing how to steer and balance at the same time. Counter-steering for fast turn entry, slow lateral responses compared with cars, and long look-ahead times for riders are consequences of motorcycle-type steering [17].

The limit behavior is substantially unknown in a scientific sense, although some anecdotal accounts suggest a very pleasant riding sensation and safety at the limit [18]. From a fundamental standpoint, one might imagine that breakaway at the front would be rare, as with motorcycles, and that the breakaway behavior at the rear would be more progressive and informative than with motorcycles. This thinking stems from the idea that reaching the adhesion limit with a sideslipping tire is the culmination of a process in which the sliding region in the tire/road contact extends further and further forward until it is all embracing. In this process, the sideslipping increases observably, and the driver is warned of the impending limit. The system dynamics are identifiable. The corresponding behavior of a motorcycle tire is less benign. Since side-force generation by camber does not involve sideslipping, reaching the adhesion limit is more sudden and less informative. Thus, it may well be that the three-wheeler of the type under discussion has advantageous limit characteristics.

Some researchers have considered it essential to the mass appeal of tilters that they respond to steering control more like a car [19–21] and, therefore, they need active tilt-control systems. To react the roll torque of the body tilt system, either a nonrolling frame is necessary or the suspension system at the two-wheeled end of the tricycle must be made active, so that the torque reaction is taken at the tire contacts with the ground. By means of tilt control, it is imagined that a much greater proportion of the population will be potential purchasers. Car-like steering can be characterized as involving steering wheel position control and lateral acceleration response at the mass center in the same sense as the steering input. If this is to be achieved with a tilter, the roll control system must be designed to yield leaning-in at the commencement of a lateral maneuver, continuing on to a sufficiently leaned-in steady turning state to substantially reduce or eliminate lateral load transfer.

During the transient, the nonrolling body on the one hand or the whole controlled vehicle on the other hand has to avoid overturning. In the case of a vehicle of the type shown in Figure 13.15, overturning is potentially a response to lateral forces developed at the nonrolling frame tires at ground level but also to those applied to it at the rolling joint and to the reaction torque to that applied to control the tilting frame. Under transient conditions, the nonrolling body is more vulnerable to overturning than a rigid-framed three-wheeler, which is quite unsatisfactory. The situation is similar in principle with an actively suspended tricycle. The more vigorously the tilt-control system deals with the steady-state load transfer problem, the more risky is the transient response to



FIGURE 13.18 The Vandenbrink Carver, showing the front suspension and steering system of the type shown in the right-hand parts of Figures 13.12 and 13.13. (From http://www.carver-technology.com)

a rapid steering input. An unequivocal account of this important difference between the transient and the steady-state behavior of a car-like tilter is given in Ref. [21], but it appears not to have been thought of in Refs. [19,20], where analysis was based on a mathematical model allowing no roll of the trolley, thereby blocking out the problem. However, even in Ref. [21], the rollover problem is not addressed fully.

General user experience with road vehicles that rely on driver restraint to avoid overturning suggests that such vehicles are dangerous and will overturn in practice to a problematic degree. Overturning is a process that includes little in the way of warning. It is difficult for drivers to learn how to avoid it, without employing totally conservative strategies, since, if one nearly overturns, little indication of near disaster is felt and the existing situation persists.

To avoid overturning both under transient and steady-state conditions, with a narrow and relatively tall vehicle, it seems that an active tilt system will have to compromise between car-like and motorcycle-like responses. A trade-off must be made between speed of response to a rapid steering control input and the overturning risk. Full and detailed simulation studies or extensive experimentation is needed to yield appropriate designs. Even then, it may well transpire that some motorcycle-like properties, like counter-steering, will have to be accommodated. It appears doubtful that active tilt control brings enough benefits to compensate for the costs, although there is no doubt that vehicles like the Carver (Figure 13.18) have a very special appeal. The Carver, which is now near to production, has a hydraulic steer by wire system and appears to respond to steering wheel displacements like a car. Running at modest speeds on gravel, it appears to be controllable when sliding both front and rear ends. A crucial test would appear to be one in which rapid steering displacements are applied in high-speed straight-running on high-friction surfaces, this being most challenging in respect of overturning.

13.3.2.2 Stability and Response Properties of the Ariel 3

The results of the straight-line stability analysis of the Ariel 3 tricycle in Ref. [15] are now revisited. Modern multibody systems software can be brought to bear on the problem. and more elaborate



FIGURE 13.19 Ariel 3 front wheel tilting tricycle produced commercially by BSA in the late 1960s and early 1970s (from http://www.mopedarmy.com). The center frame tilts around a slightly inclined longitudinal joint with the rear frame, restrained by a torsion bar, stiff enough that the tricycle remains upright with no rider aboard. The origins of the Ariel 3 are discussed in. (From Jackson, S., The Ariel 3 was classified as an articulated three wheeled moped, http://www.users.globalnet.co.uk.)

models than were previously practical can be built without particular difficulty. The model of [15] has been rebuilt using AutoSim [11], with the modest elaboration that the tires are now radially compliant and the turn-slip terms in the tire forces and moments have been omitted. All the parameter values describing the Ariel 3 tricycle can be found in the original paper. The machine itself is shown in Figure 13.19.

Characteristic root real parts are shown as functions of speed, over the range of 0.5–32 m/s in Figure 13.20, for variations in the stiffness of the roll-restraining torsion bar. Cases treated involve no torsion bar, the standard bar, and a torsion bar with twice the nominal stiffness. These stability properties are similar to those of motorcycles, with the roll instability apparent at low speeds, self-stabilization at about 5 m/s, and weave and wobble oscillatory modes. The stiffness of the roll-restraining torsion bar influences the rider's balancing problem at speeds above 5 m/s, but it is unlikely that a rider would sense much difference between the cases, since the balancing is not difficult in any of them.

Some responses to a sinusoidal steering torque input with magnitude 3.2 Nm and circular frequency 2 rad/s when traveling at 20 m/s are shown in Figures 13.21 through 13.23. The maneuver is somewhat arbitrary, being chosen to show the buildup of large responses over several seconds.

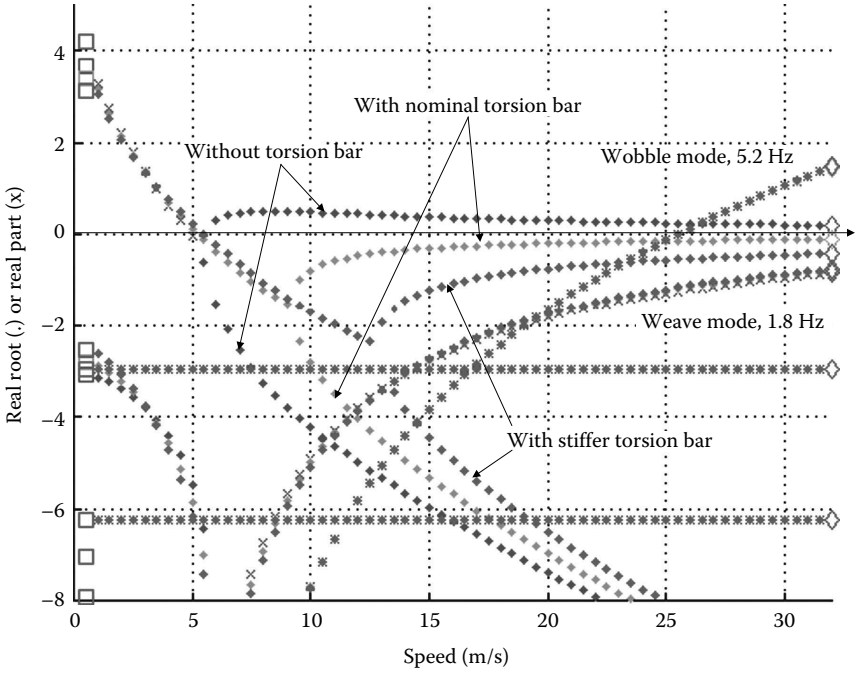


FIGURE 13.20 Ariel 3 real parts of roots as functions of speed without torsion bar, with nominal torsion bar, and with a torsion bar with twice the nominal stiffness. The tricycle can be seen to have weave and wobble modes, with the wobble mode becoming unstable at about 25 m/s speed in all three cases. The weave mode is also substantially uninfluenced by the torsion bar, but the torsion bar does stabilize the capsize motion above about 5 m/s.

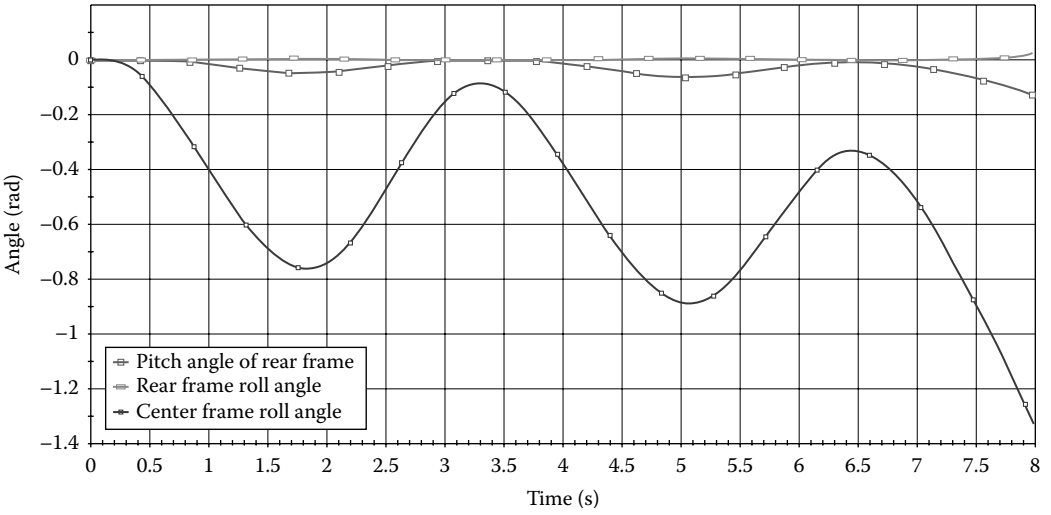


FIGURE 13.21 Ariel 3 pitch and roll angle responses to a sinusoidal steering torque input when traveling straight at 20 m/s.

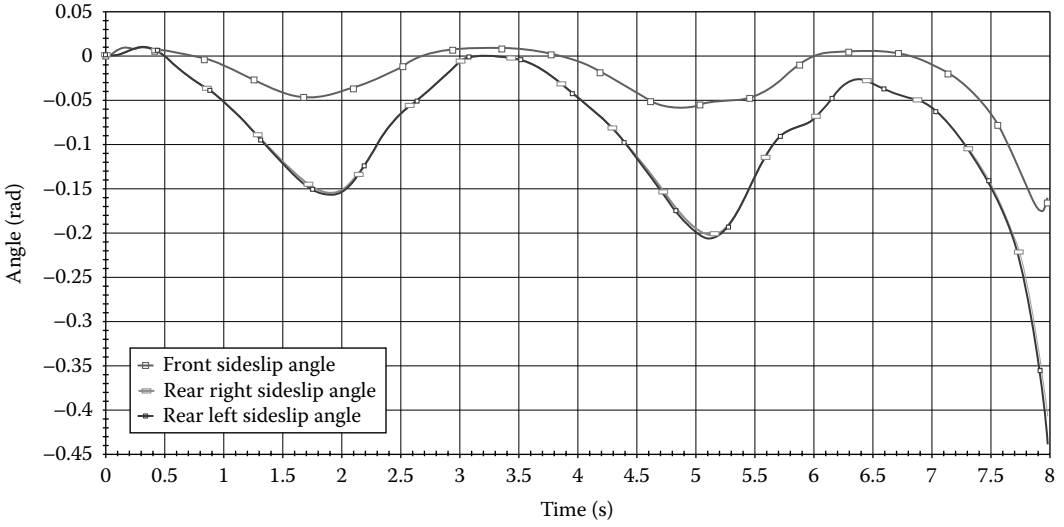


FIGURE 13.22 Ariel 3 sideslip angle responses to a sinusoidal steering torque input when traveling straight at 20 m/s.

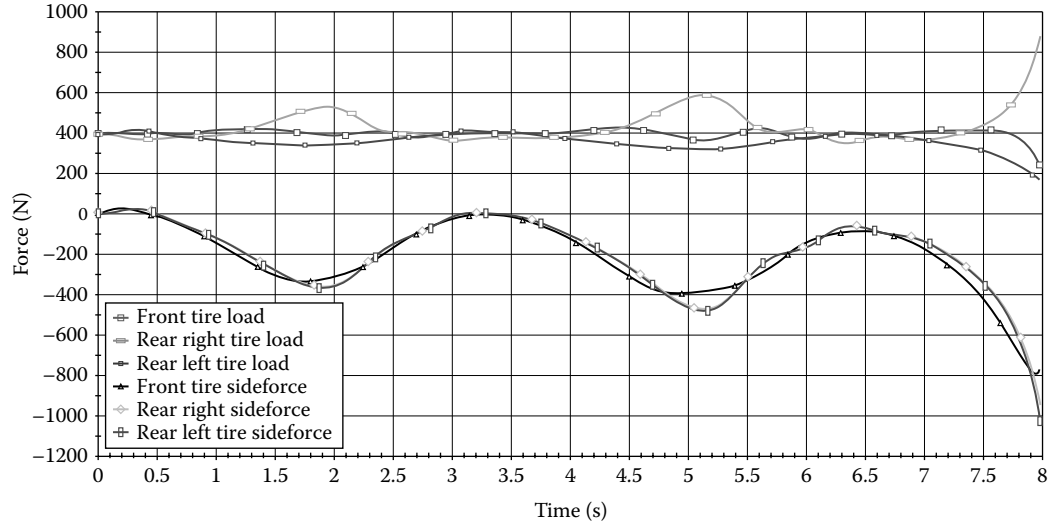


FIGURE 13.23 Ariel 3 tire load and side-force responses to a sinusoidal steering torque input when traveling straight at 20 m/s.

The responses show the buildup of a center frame roll angle larger than 1 rad, while the rear frame roll angle remains small and the load transfers are quite modest. Near to the end of the run, the rear tire sideslip angles grow very large, with that at the front about half as big.

13.4 CONCLUSIONS

How two-wheeled vehicles operate and what matters in their design have been discussed. Types of motorcycle frames, suspension and steering systems, and the antidive and antisquat properties, in particular, have been described. Contrasting types of front suspension and steering systems have

been examined, with their strengths and weaknesses highlighted. A brief consideration of nontilting three-wheeled vehicles is followed by a more detailed account of tilters. Special attention is paid to rollover. This is not a problem with three-wheelers that are controlled in the manner of a motorcycle with the relatively slow responses of a bank-to-turn machine, but it is crucial that it is dealt with effectively in the case of a tricycle intended to steer like a car. A small number of free-tilters have been made commercially, among them the Ariel 3. A new approach in the form of the Carver is near to commercial manufacture. The Carver has an active tilt-control system and appears to steer like a car while banking like a motorcycle. The limit behavior has certainly been tested extensively at moderate speeds on loose surfaces, as evidenced by video material that can be seen on the Carver website.

Tilting tricycles have been analyzed to only a very small extent, with some analyses spoiled by the oversimplified models employed. It is possible to build realistic models of such systems using contemporary multibody software, and the point is illustrated by revisiting an analysis of the Ariel 3 straight-line stability properties, with additional limited results on responses to steer torque control. Load transfers across the rear axle are shown to be small until extreme roll angles of the center frame occur. Controllability of the limit behavior is a vital issue, in respect of which scientific data are lacking. There seems to be an opportunity to apply "worst-case methodology" using optimal control theory [22], to determine the set of initial conditions and control inputs posing the greatest challenge to the rollover immunity.

REFERENCES

1. Sharp, R. S., Stability, control and steering responses of motorcycles, *Vehicle System Dynamics*, **35**(4–5), 2001, 291–318.
2. Schwarz, R., Accident avoidance capabilities of unconventional motorcycle configurations, SAE Paper, 790258, 1978.
3. Aström, K. J., Klein, R. E., and Lennartsson, A., Bicycle dynamics and control: Adapted bicycles for education and research, *IEEE Control Systems Magazine*, **25**, 2005, 26–47.
4. Foale, T., *Motorcycle Handling and Chassis Design—The Art and Science*, Tony Foale Designs, Alicante, Spain, 2003 (http://www.tonyfoale.com).
5. Cossalter, V., *Motorcycle Dynamics*, Race Dynamics, Greendale, WI, 2002.
6. Sharp, R. S. and Giles C. G., Motorcycle front wheel patter in heavy braking, In *The Dynamics of Vehicles on Roads and on Railway Tracks*, J. Karl Hedrick (ed.), *Proceedings of 8th IAVSD Symposium*, Swets and Zeitlinger B. V., Lisse, Western Netherlands, 1983, pp. 578–590.
7. Cameron, K., *Sportbike Performance Handbook*, MBI Publishing, Osceola WI, 1998.
8. Sharp, R. S., The influence of frame flexibility on the lateral stability of motorcycles, *Proceedings of the Institution of Mechanical Engineers Part C*, *Journal of Mechanical Engineering Science*, **16**(2), 1974, 117–120.
9. Wilson, H., *The Encyclopedia of the Motorcycle*, Dorling Kindersley, London, U.K., 1995.
10. Sharp, R. S., Evangelou, S., and Limebeer, D. J. N., Advances in the modelling of motorcycle dynamics, *Multibody System Dynamics*, **12**(3), 2004, 251–283.
11. Sharp, R. S., Evangelou, S., and Limebeer, D. J. N., Multibody aspects of motorcycle modelling with special reference to AutoSim, In *Advances in Computational Multibody Systems*, J. Ambrósio (ed.), Springer-Verlag, Dordrecht, the Netherlands, 2005, pp. 45–68.
12. Kane, T. R. and Levinson, D. A., *Dynamics: Theory and Applications*, McGraw-Hill, New York, 1985.
13. Sharp, R. S., Influences of suspension kinematics on pitching dynamics of cars in longitudinal manoeuvring, In *The Dynamics of Vehicles on Roads and on Tracks*, R. Fröhling (ed.), Supplement to VSD **33**, Swets and Zeitlinger B. V., Lisse, Western Netherlands, 2000, pp. 23–36.
14. Sharp, R. S., Variable geometry active rear suspension for motorcycles, *Proceedings of AVEC 2000, 5th International Symposium on Advanced Vehicle Control*, Ann Arbor, MI, August 22–24, 2000, pp. 389–395.
15. Sharp, R. S., The stability and control of pivot framed tricycles, In *The Dynamics of Vehicles on Roads and Tracks*, J. Karl Hedrick (ed.), Swets and Zeitlinger B. V., Lisse, Western Netherlands, 1984, pp. 564–577.

16. Foale, T., *Some Technical Aspects of Tilting Trikes*, Tony Foale Designs, Alicante, Spain, 1999, pp. 24 (http://www.tonyfoale.com).
17. Sharp, R. S., Optimal linear time-invariant preview steering control for motorcycles, In *The Dynamics of Vehicles on Roads and on Tracks*, S. Bruni and G. Mastinu (eds.), Supplement to VSD **44**, Taylor & Francis, London, U.K., 2006, pp. 329–340.
18. Jackson, S., The Ariel 3 was classified as an articulated three wheeled moped, http://www.3wheelers.com. Accessed July 2013.
19. Hibbard, R. and Karnopp, D., Twenty first century transportation system solutions—A new type of small, relatively narrow active tilting commuter vehicle, *Vehicle System Dynamics*, **25**(5), 1996, 321–347.
20. So, S.-G. and Karnopp, D., Active dual mode tilt control for narrow ground vehicles, *Vehicle System Dynamics*, **27**(1), 1997, 19–36.
21. Snell, A., An active roll moment control strategy for narrow tilting commuter vehicles, *Vehicle System Dynamics*, **29**(5), 1998, 277–307.
22. Ma, W. and Peng, H., Worst-case vehicle evaluation methodology—Examples on truck rollover/jack-knifing and active yaw control systems, *Proceedings of AVEC 1998, 4th International Symposium on Advanced Vehicle Control*, SAE of Japan, Tokyo, Japan, 1998, pp. 299–305.

14 Race Cars
Frame, Suspension, Aerodynamics

Andrea Toso

CONTENTS

14.1 Introduction 497
14.2 Vehicle Dynamics for Motor Racing 498
14.3 Racing Car Concept 500
14.3.1 Metrics 500
14.3.1.1 Functions 501
14.3.1.2 Weight 502
14.3.1.3 Time 502
14.3.1.4 Cost 503
14.4 Racing Car Key Components 504
14.4.1 "Frame" 504
14.4.2 Suspensions 505
14.4.2.1 Contact Patch Forces 505
14.4.2.2 Steering and Driver Feedback 505
14.4.2.3 Suspension Tuning 506
14.4.3 Aerodynamics 506
14.4.3.1 Aerodynamic Loads 506
14.4.3.2 Pitch Sensitivity 508
14.4.3.3 Polar Curve and Aerodynamic Tuning 508
14.5 Racing Car Performance: Lap Time 509
14.5.1 Car Performance 510
14.5.2 Car Development 510
14.5.2.1 Setup 510
14.5.2.2 Tires 511
14.5.2.3 Driver 512
14.5.2.4 Driving Simulator for Race and High Performance Cars 513
14.6 Conclusions 514
References 514

14.1 INTRODUCTION

The sport of motor racing is heavily constrained by the money and the time to design, build, and develop a car ready for the race events.

Winning in the sport of motor racing requires not only to exploit, to the very maximum extent, the rules that define tires, transmission, engine, and aerodynamics but also to achieve the most efficient driver/vehicle combination for the actual race conditions.

In every motor racing series, the driver/vehicle system operates under stressful conditions where the overall performance is affected by decisions taken before and during the events, usually with

insufficient time or data [1]: In fact, the sport of motor racing occurs within a poorly controllable environment (ambient and track surface conditions, traffic situations), and the final result depends on the interaction between the driver and the vehicle, on how well physically trained the driver is to sustain the various and prolonged physical efforts, and how mentally strong the driver is to consistently control the vehicle [2–4].

Motor racing covers many diverse forms of competitions (endurance, sprint, rally, hill climb), of tracks (permanent or temporary circuits), of vehicle layouts (front-, rear-, or four-wheel drive), and of vehicles (cars, motorcycles, motorboats, quads, etc.) and requires a fair amount of qualified technical personnel (mechanics, engineers, designers, etc.) and also skilled people in the administration, logistics, managing, etc.; all the aforementioned people contribute to a winning car [5].

Some subjects of road car vehicle dynamics are of limited or even marginal significance in motor racing, like driver comfort requirements and active suspensions, while others are generally banned as illegal, like ABS and ESP, since they diminish the importance of the driver on the final result [6].

The following section essentially refers to rear-engine four-wheel passive suspension road racing cars competing in closed circuits, with the traction available only at the rear axle and the steering limited to the front axle.

14.2 VEHICLE DYNAMICS FOR MOTOR RACING

The dynamics of the car body is similar to the dynamics of the aircraft body. In the past, some cross-fertilization between the aircraft dynamics and the ground vehicle dynamics has given fruitful results (see [2,7–9]), even if aircraft dynamics does not deal with the highly nonlinear behavior of the tires (Figure 14.1).

Vehicle dynamics for motor racing (VDMR) is an engineering discipline that studies the interdependencies between driveline, aerodynamics, chassis, suspensions, and tires by means of simulation techniques and verifications on the field, most of the times the actual race

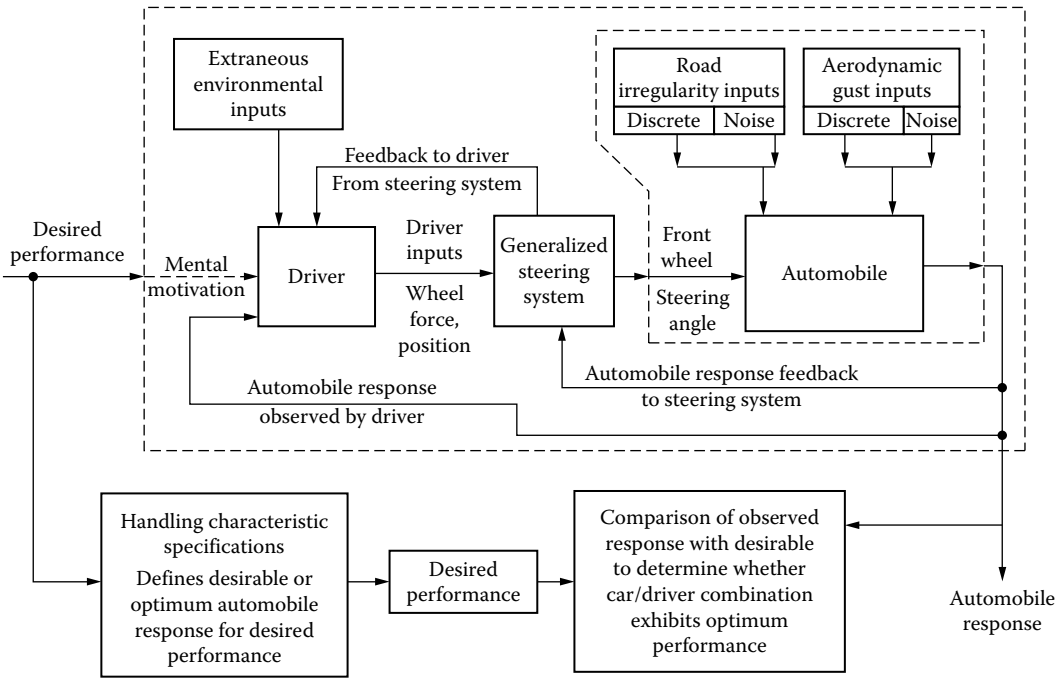


FIGURE 14.1 Car/driver combination: generalized block diagram.

tracks, with the ultimate goal to minimize the lap time, with adequate handling stability and controllability, and with acceptable ride qualities [10].

Some typical basic competences required to the VDMR engineer are straight line and cornering performance simulation, aerodynamics force optimization, tire modeling, onboard-recorded track data correlation with simulated maneuvers, analysis of the driver-to-vehicle interaction, and derivation of the loads onto the main chassis structures and components [1,11,12].

VDMR therefore provides a solid foundation for the vehicle initial concept and guides the ongoing development imposed by the challenges raised by other competitors or by new available technology: in this respect, VDMR is of strategic relevance alongside the vehicle structural and mechanical design, the aerodynamic research, and the manufacturing process.

At a higher complexity level, VDMR also focuses on enabling the racing driver to race consistently, lap after lap, with the highest possible acceleration in the appropriate directions. Accelerations are the direct consequences of forces and moments produced by tires and aerodynamic field. The driver/vehicle system able to consistently generate the highest forces generally wins.

In a reference system fixed with the moving vehicle, accelerations are not limited to the longitudinal axis (velocity modulus variation in braking and acceleration phases) but are also generated along the transverse axis (due to either a modulus or a direction variation of the vehicle speed) and along the vertical axis (due to wheel load variations produced by the road irregularities and in general by forces and moments generated by any means) [2].

Maximum acceleration out of a corner, maximum deceleration while braking into a corner, maximum lateral acceleration in a corner, maximum combined longitudinal and lateral accelerations during the entire corner, and minimum vertical wheel accelerations over the road irregularities are, therefore, fundamental to achieve the optimum performance with the nonobvious consequence that, in motor racing, the car velocity should vary as quickly as possible. Exceptions to the goal of always maximizing the accelerations might occur in temporary or intentional situations like endurance race events, traffic situations, safety and reliability concerns, and intended fuel strategy.

It is common practice for the vehicle dynamics engineer to correlate the longitudinal and the lateral accelerations, either simulated or recorded, so as to determine the maximum combined car acceleration envelope, or g–g diagram, at different speeds or for different cornering radii and to study ways to further widen it [13,14].

One of the biggest challenges for the VDMR expert in correlating the simulation models with the real vehicle is that the acceleration envelope depends on the current environment and track conditions, on many and often nonlinear suspension parameters, on engine characteristics, on the always changing overall weight (due to fuel consumption), on many nonlinear aerodynamics settings, on hardly repeatable driver inputs, and on limited useful life of some fundamental components like the engine, the racing tires, the transmission gears, the wheel bearings, and the suspension ball joint.

The focus of VDMR is not only limited to widening the acceleration envelope but also to offering an adequate level of controllability and stability near the limit of the acceleration envelope. The handling parameters have to suit each individual driver like a “haute-couture” handmade dressing to let him or her consistently drive, race, and win: The same racing car cannot be driven in the same way by two different drivers because of the different physical and mental preferences in terms of inherent vehicle stability, gain, and response time to the driver controls, like brakes, throttle, gearshift, and steering. These quite individual characteristics require, at a high level of vehicle dynamics analysis, a deep knowledge of the driver as a unique performer and thus involve psychology, ergonomics, and neurology competences.

Ride performances (like driver comfort [see Chapter 36], sprung mass accelerations, suspension wheel travel, engine vibration insulation, aerodynamic noise) are not generally a priority in motor racing [15].

Since maximum chassis torsional stiffness and minimum suspension compliance (due to joint stiffness, mounting bracket stiffness, wheel bearing stiffness, wishbone deformation, etc.) are needed to achieve the highest efficiency in applying the forces at the four wheels, the racing car has

all the mechanical parts mounted rigidly, without ride-compliant or vibration-insulation devices, either to assemble the suspension elements or to support the engine and the driveline [16].

Ride quality comes into consideration only if its level is so poor as to endanger the overall lap time performance. If, for example, over curbs or over a local track surface irregularity, the vertical forces at the four wheels vary too much, then the longitudinal and the lateral forces produced inevitably affect the car's handling. To overcome these local ride problems, springs and dampers are tuned for the specific race track but always with particular attention to the handling performance over the entire lap [17].

To study these particular problems and to improve the ride quality performance above an acceptable level, tests are carried out at dedicated indoor test rigs, where four, seven, or even more independent actuators simultaneously act on the car and reproduce any generic road irregularity and even any specific race track profile [18].

14.3 RACING CAR CONCEPT

Racing cars are special vehicles to be driven in competition only; therefore, dedicated design, manufacturing, test, and maintenance techniques are required.

The design cannot always be verified on the field and for the actual conditions of use due to the often rule-limited available track time; moreover, racing cars are produced in limited quantities (8 cars/year for Formula 1, 60 cars/year for a Formula 3), are expected to race for limited time (one or two seasons), are assumed to be continuously developed to match fast-evolving rules and competitors' performance, and are required to be easily tuned for the specific environment or race track conditions.

The frame, or monocoque or tub as it is sometimes referred to, is the most important structure of the racing car. The front suspension and the engine, generally a stressed member, are directly bolted on.

The frame also incorporates the fuel compartment and the survival cell, where the driver is seated. The transmission is mounted on the rear face of the engine, and the rear suspension is bolted on the transmission. The frame, the engine, and the transmission define the spine that, together with the driver, accounts for almost all the sprung mass and where all the remaining auxiliary systems, such as bodywork, aerodynamic devices (if allowed), cooling elements, safety devices, and electronics, are attached to (Figure 14.2).

14.3.1 Metrics

Performance in motor racing, for a complete vehicle and for its single components, is evaluated with few metrics, essentially weight, cost, time, and functions; examples of component functions are the capability of aerodynamic parts to generate downforce or drag, the ability of suspension elements to manage the wheel load variations during the maneuvers and to transmit torque and loads between the tire and the ground, stiffness or strength of structural parts, cooling efficiency,

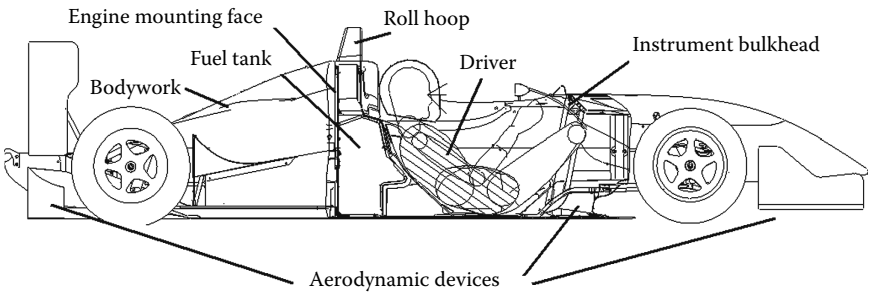


FIGURE 14.2 Racing car layout.

TABLE 14.1
Main Racing and High-Performance Car Characteristics

	Race	Road
Dry mass (no driver) (kg)	450–750	1300–1800
Wheelbase (mm)	2700–3000	2500–2700
Yaw inertia (kg m^2)	900–1150	1700–2500
Downforce (@50 m/s)/weight (%)	100–200	3–5
Power (kW)	150–550	150–500
Power/mass (kW/kg)	0.30–0.80	0.15–0.30
Tire adherence coefficient	1.3–1.7	1.0–1.1
Max acceleration (lateral/longitudinal) (g)	2.5–4.5	1.1–1.3
Center of gravity height (mm)	225–275	350–450

brake distance, steering effort with different steering wheel angles, engine available torque, easy access for tuning and replacement, etc.

When compared to a road car, the racing car has much fewer parameters to be optimized, but these parameters have to be really excellent; the limited set of performance metrics inevitably determines a simple and extremely efficient layout.

The typical wheelbase of racing cars is higher than the one for the high-performance road car, but the weight and the yaw inertia are much lower. So the racing car is subjected to much stronger external (aerodynamics) and internal (braking, engine) forces and moments, and the consequent accelerations are much higher.

A comparison between racing and high-performance road car basic parameters is given in Table 14.1.

14.3.1.1 Functions

The structural function is one of the preliminary ones for all racing components: stiffness is a factor for heavily stressed structural components, like suspension parts and aerodynamic devices, up to the point that sometimes stiffness is more important than ultimate strength. Stiffness versus strength is the typical dilemma in motor racing, not only for the designer, but also for the vehicle dynamics expert, since higher stiffness generally requires more weight or increased cost, if more exotic materials are used.

The most important performance functions of a racing car, as a complete vehicle, are the reliability, the ease of adjustment and replacement, and, above all, the lap time.

A useless part can fail and must not be mounted on any racing car and a necessary part has to fail shortly after the required life.

The racing car must be reliable enough to consistently finish the race and each component has to be designed and manufactured for the appropriate life (and not much more). A part with no evidence of pitting, fatigue, plastic deformation, corrosion, superficial damage, etc., at the end of the targeted life cycle is too heavy and potentially subjected to a redesign. The life cycle can be one race only (like in Formula 1 or in sports car for the 24 h race of Le Mans) or one complete season for a commercially available racing car like a Formula 3.

Redesign can allow for a lighter component, with positive drawbacks on the weight budget, but can carry negative consequences on the money budget and on the production timing.

The racing car must also be easily adjustable, and the design must account for the adequate tuning parameters in order to cover all the potential needs, in terms of cooling, cockpit controls, suspension and aerodynamic settings, etc.

The design must also account for frequent installation and removal of parts, like the engine, the gears, the dampers, and the coolers as the time available at the race track is limited.

For example, valuable track time can be gained if a spring change requires less time: actually, lap time performance is related not only to the car suspension and/or aerodynamic configuration but also to the driver's confidence in the car's expected reactions to the control inputs and to the detailed and fresh memory of the most lap time-efficient driving trajectory.

Lap time is by far the most important performance index for any racing car and it will be covered in a dedicated section.

14.3.1.2 Weight

There is no useless piece in a racing car!

The racing car must be light enough not to exceed the minimum weight limit imposed by the rules but also strong enough not to cause the failure of any component under racing use.

Weight is a primary consideration for the overall performance of any racing car. An overweight car produces less accelerations in all directions for the same amount of generated forces. The weight budget is carefully allocated and continuously adjusted among the different systems and components during the design and development phases: it is the responsibility of the VDMR expert to help the designer allocate the weight budget and to decide over the conflicts between reliability, weight, costs, performance, production time, etc.

To emphasize the characteristics of short transient response, the VDMR and the designer always try to move the heavy components (engine, driver, fuel, transmission, brakes) as close as possible to the center of gravity where even a few millimeters do make a noticeable difference for the expert driver.

Another responsibility for the VDMR expert is the optimal location of the center of gravity between the two axles. The longitudinal location is driven by the tire adhesion and wear properties, the extent of aerodynamic forces and their distribution between the front and the rear axle, the traction torque available at the rear wheels, and the specific race track (i.e., a twisty and tight corner complex or a high-speed speedway).

The design and test efforts to continuously reduce the weight of the components are justified by the quest of lowering the vertical location of the center of gravity to the point that sometimes the accessibility to the components and to the auxiliary systems (like suspension elements, engine boxes, and electric looms) lose priority. If the car is reliably designed underweight, the available ballast, typically between 3% and 10% of the overall car weight and function of the time and money budget limitations, can be secured underneath the car.

A still open question refers on how the inertia tensor can be optimized by adding ballast.

Since lap time performance changes with the overall car weight, the fuel consumption and refueling strategy in the pits are additional aspects studied by the VDMR expert.

14.3.1.3 Time

The time metric considers the interval between the design concept and the race-ready conditions. A car able to compete in an event, but not ready at its maximum performance potential, is the evidence of a big time-planning mistake.

The time schedule to design and build a complete racing car is very tight, and the design, production, and development phases are pretty much overlapped over each other; careful management of all critical aspects like frequent data exchanges between the departments, possible conflicts of responsibilities, and stressed interpersonal relationships are essential (Figure 14.3).

The design review and the test development phases are the most important ones, and multiple iterations are needed up to the point where it becomes impossible to separate the two. Production is always in the loop as new development parts have to be designed in a very short time, be quickly manufactured with high quality and in small quantities, and be tested up until the last minute before a race event. Sometimes, this cycle does not stop even after the race if the results have not been satisfactory, provided that the rules or the budget allows.

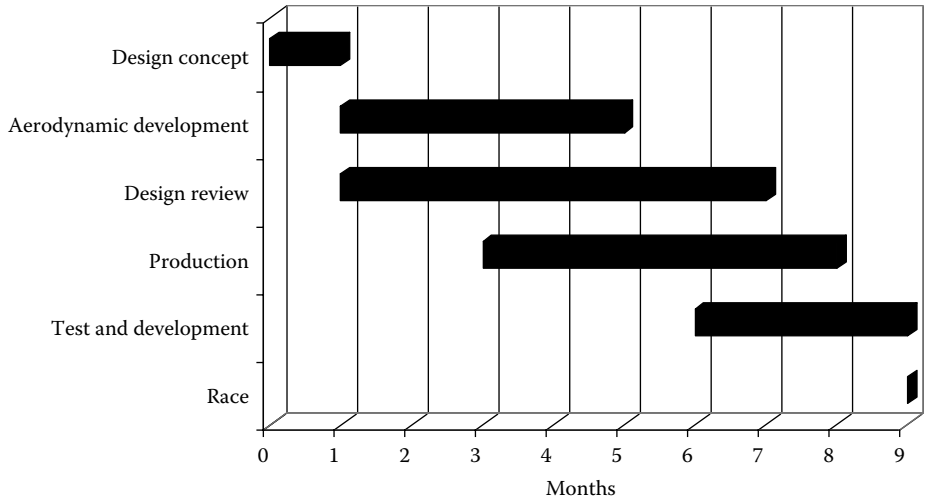


FIGURE 14.3 Time frame to design/build and run a racing car (representative of a Formula 3 project).

The previous aspects very well highlight why a racing car is so different from a road production car. For the racing car, the prototype and the final product do coincide, the product engineering phase is essentially the product development phase, the design is a continuous design review, the time to market is a matter of a few months, and the production costs are marginal if compared to the design, research and development, and testing ones.

Here, the VDMR expert has a pivotal role in the never-ending cycle "design–production–test and development" to objectively predict the lap time reduction per reduced weight and per increased cost or time allocated in the design, production, and test phases.

14.3.1.4 Cost

The racing car as a global vehicle, and every component of it in detail, must be essential and simple in order to maintain the manufacturing time and costs under control.

One piece with multiple functions is usually lighter and cheaper than two pieces with separate functions.

Due to the low production volume and to the identity between the prototype and the actual race car, the production costs are generally lower than the design, test, and development costs.

Attention to costs is always important in all forms of racing competition, since the money budget for any motor racing organization, whether a racing team or a chassis manufacturer, has direct and immediate consequences on the overall car performance (Table 14.2).

TABLE 14.2
Main Racing Car Characteristics for Three Single-Seater Racing Cars (2005)

	a	*b*	*c*
Budget (euro × 1000 as June 2006)	300	800	3,000
Involved people	6	12	40
Complete car mass (no driver/no fuel) (kg)	475	600	710
Engine power (kW)	160	315	510
Mileage per season (km)	12,000	10,000	15,000
Running cost (euro/km)	25	80	200

14.4 RACING CAR KEY COMPONENTS

The three most distinctive subsystems of a racing car, if compared to a road car, are the frame, the suspensions, and the aerodynamics.

14.4.1 "Frame"

The frame of a racing car has two main functions, namely, safety and stiffness. (For additional details on frames, see Chapter 26.)

The frame, or at least a part of it, has to ensure adequate protection both to the driver and to the fuel tank.

The rules governing the racing car design clearly define a set of required tests to assess the frame static stiffness in crucial areas like the fuel tank, the instrument and pedal bulkheads, the ultimate strength in the roll hoop area, and the dynamic strength to survive in severe frontal and lateral crash tests [6].

Provided that the frame does resist all the safety tests, the structure has to be as stiff as possible in torsion, in order to allow the appropriate roll stiffness setting between the front and the rear axle [19,20].

An increase in the frame torsional stiffness is always a positive factor, but it has to be achieved with the minimum weight increase. For this reason, when comparing different frames, or different design alternatives, the frame torsional stiffness is usually normalized to the frame mass (see Table 14.3).

The search for the optimum frame stiffness versus weight has driven the racing car manufacturers to develop composite construction structures, derived from the military and the aeronautic industries. The anisotropic mechanical properties of the composite fibers, like carbon, Kevlar™, and fiberglass (pasted by means of resins, like epoxy and phenolic ones), can be exploited with great advantage to increase stiffness and reduce weight, but as long as the load path distribution is known within an acceptable accuracy. The low production volume of racing cars and the high cost of the raw material and the manufacturing process do limit the applications only to professional racing car series [21].

The frame is only one of the three elements that constitute the "spine;" the other two, in series with the frame, are the engine and the transmission casing, when mounted as stressed members.

The frame stiffness or, in a broader sense, the "spine" stiffness is not the total wheel-to-wheel stiffness, which is the determining factor in defining the four vertical wheel loads in every load conditions, either steady state or transient: here, the missing essential contributor to the overall wheel-to-wheel stiffness is the suspension compliance, due to the wheel bearings, the suspension ball joints, the suspension mountings, the suspension arms, the bell-crank (or rocker) post, etc. [19].

TABLE 14.3
Frame Weight and Stiffness

	a	*b*	*c*
Complete car mass (no driver/no fuel) (kg)	475	600	710
Frame mass (kg)	45	58	67
Engine mass (kg)	105	130	145
Transmission mass (kg)	55	75	80
Frame torsional stiffness (N m/rad)	15,000	22,000	26,000
Frame torsional stiffness/frame mass ($m^2/s^2/rad$)	333.3	379.3	388.0
Frame mass/car mass	0.095	0.097	0.094
Spine mass/car mass	0.43	0.44	0.41
Wheel-to-wheel torsional stiffness (N m/rad)	7,850	9,300	11,300

Wheel-to-wheel torsional stiffness is usually defined and measured without taking into account the tire stiffness.

Since the designer and the VDMR expert have little influence on the engine or driveline torsional stiffness, the frame weight, cost, and stiffness are of main interest.

14.4.2 Suspensions

As for any four-wheeled vehicle, the distribution of the contact loads between the wheels and the ground is not statically determined. (For additional details on suspensions, see Chapter 21.)

14.4.2.1 Contact Patch Forces

Suspension systems for racing cars are designed, developed, and tuned primarily to manage the wheel force distribution between the four wheels during all the maneuvers and to minimize the wheel force variation at the contact patches over the road irregularities. The problem has a difficult solution because the optimum suspension parameter settings are variable with the particular maneuver, driver's input, road irregularities, car speed, and instantaneous accelerations. The vertical wheel forces as well as the lateral and the longitudinal forces at the ground depend on the inertia of the sprung and unsprung masses and on the suspension stiffness and damping.

Every variation of the vertical wheel load at the contact patch affects the tire ability to generate grip, the tire instantaneous working temperature, and the tire wear.

Since tire grip, temperature, and wear have to be balanced front to rear and since every suspension adjustment at one wheel corner does influence the loads at the same wheel corner and at all the other three wheel corners, the quest of the correct balance at the highest possible level of grip requires very sophisticated modeling techniques, continuous cross validations with the real vehicle, and detailed indoor and outdoor measurements of the many and highly nonlinear suspension parameters like vertical and roll stiffness, damping, suspension compliance, friction, backlash, and hysteresis, some of them variable with the working temperature and wear.

The spring stiffness selection is essentially based on the target sprung mass resonance frequencies at the front and at the rear axle, which is a function of the tire stiffness, unsprung mass, relevance of the aerodynamic loads, and aerodynamic load variations with front and rear ride height (RH) (see later Section 3.3), vertical suspension compliance, etc.; other factors coming into play are the engine to power ratio and the race track typology that might have a preponderance of heavy braking or hard acceleration or twisty corners or bumpy sections.

The roll stiffness and its distribution between the front and the rear axles is determined by the race track characteristics, the typical lateral accelerations and cornering speeds, the center of gravity height to the track width ratio, and the tire characteristics in the terminal nonlinear range of slip and camber angles.

The damping characteristic selection follows the spring and the roll stiffness choice, which is responsible for defining the RH in the straights and the roll angle in the center of the corner. Dampers, in fact, are effective and powerful devices to tune the handling of the car during the transient maneuvers, where the wheels have a significant inherent velocity relative to the unsprung mass.

14.4.2.2 Steering and Driver Feedback

The second function of the suspension system of racing cars is to provide the driver with the correct feeling, in terms of vehicle response time, vehicle response gains, steering effort, etc., so the driver can "read" the car at the limit of performance and set the desired trajectory.

The human brain tends to prefer a proportional input–output relationship to be able to subconsciously predict the response to a different input; similarly, a racing driver appreciates when the car reactions to the control inputs are linear. Consequently, all suspension nonlinearities, like motion ratios, roll and ride stiffness, bump rubbers, and spring preloads, have to be carefully considered as they reduce the car predictability.

TABLE 14.4
Steering Ratio (Steering Wheel/Front Wheel)

	a	*b*	*c*
Steering ratio	11.0–14.0	12.0–15.0	12.5–16.0

The driver controls the racing car in a complex way, partially in open-loop control mode and partially in feedback control mode. Because of the reduction of the yaw damping with speed, the higher the vehicle speed, the more the driver has to drive in an open-loop mode. For racing on ovals, for example, the speed is very high, up to 240 mph (over 100 m/s): There is no time to finely readjust the control inputs, and it is necessary to apply the precise steering input to have the vehicle take a set on the intended trajectory at the end of the turn, seconds later, away from the inside and outside walls [2,11].

On the very opposite, for rally races, where the speed is much lower, the vehicle is driven almost in closed loop, and an instant feel of the car attitude (yaw, sideslip, and roll) is needed for the driver to control the car on a highly variable grip surface.

The aforementioned two extremes of driving a race car clearly require different suspension settings not only to maximize, at every instant, the contact between each tire and the ground and to optimize the distribution of the vertical loads between the four wheels but also to give the driver the correct feedback information in the appropriate time, neither too late or too soon, depending on the vehicle speed and weight, engine power, race track geometry, tire properties, etc. (Table 14.4).

Cars running on oval tracks at very high speed and for prolonged time do require "slower" steering ratios, both to reduce the steering effort for the driver and to provide a higher angular resolution at the steering wheel.

14.4.2.3 Suspension Tuning

Tuning the car's handling by adjusting the vehicle reaction times and output overshoot is of primary relevance in racing where the steady-state conditions are such a marginal percentage of the overall lap time. The understeer/oversteer characteristics are not exhaustive enough to judge on the handling performance [2,22]. Tuning a racing car means changing its handling balance, or in other terms redistributing the produced grip between the front and the rear tires.

The racing tires have to be exploited at or near the maximum of their adherence capabilities, which happens in their nonlinear range of operation. The efficiency of the settings of springs, dampers, antiroll bars, and roll and pitch centers is to be evaluated away from the linear range of the tire working characteristics. Due to the nonlinear nature of the tire characteristics, the amount of lateral grip available on one axle depends on the lateral load transfer between the inner and the outer wheel and on the distribution of the roll stiffness between the two axles; therefore, the management of the vertical wheel loads is the key to tune the racing car [23].

14.4.3 Aerodynamics

14.4.3.1 Aerodynamic Loads

Aerodynamic forces acting on racing cars are roughly proportional to the square of the car speed and are comparable with the vehicle weight [24,25]. (For additional details on aerodynamics, see Chapter 16.) The main focus of vehicle dynamics when dealing with aerodynamic forces is to handle their violent variation with speed and their distribution between the front and the rear axles.

The aerodynamic forces are usually decomposed in three terms, either total downforce, drag, and total downforce balance front to rear or front axle downforce, rear axle downforce, and drag. Aerodynamic efficiency is defined as the ratio between the total downforce and the drag and comparison between two different aerodynamic configurations is not a straightforward operation since, given the same efficiency, the aero balance or the drag might be different.

TABLE 14.5
Total Aerodynamic Loads at (50 m/s), Air Density at 18°C, 0% Relative Humidity, Atmospheric Pressure 1013 mb

	a	*b*	*c*
Complete car mass (no driver/no fuel) (kg)	475	600	710
Downforce (N)	3500	5800	6900
Aero balance (%front)	42	40	38
Drag (N)	1550	1650	2550

Some aerodynamic loads, for the three representative vehicles mentioned earlier, are listed in Table 14.5.

The indirect but important effect of the aerodynamic forces on handling comes through the tires. Provided that the tires are able to generate grip (lateral or longitudinal or both combined), based on the vertical load applied, an increase in vertical load improves the capability of the tire to generate grip. Aerodynamics is a very powerful and delicate tool to achieve handling characteristics that vary with speed [26].

At a given speed, a racing car running in close proximity to the ground produces forces primarily dependent on the front and rear RH as measured at the wheel axle. This concept is referred with the term "aeromap" that is the tabular or graphical representation of the variation of how the aerodynamic parameters like front downforce, rear downforce, total downforce, aero balance, and aerodynamic efficiency and drag vary with front and rear RH (Table 14.6).

Aerodynamic forces also depend on other parameters like the heading of the car to the wind (determined by the car yaw angle), the roll angle (which primarily influences the aerodynamic loads generated by the underside of the car), and the steering angle of the front wheels (particularly important for open wheel racing cars). In the last example, the objective is to minimize the aero-balance variation due to the steering wheel angle or to achieve a desired relationship so as to modify the handling qualities for specific speeds or cornering radii during turn-in, mid-corner, power-on application maneuvers. This analysis requires a highly nonlinear simulation model because the required steering angle is function of the grip generated by the tires that

TABLE 14.6
Front Downforce Load (N) at 180 km/h (50 m/s), Air Density at 18°C, 0% Humidity, Atmospheric Pressure 1013 mb

	Front Downforce (*N*)
Front RH = 10 mm/rear RH = 15 mm	2800
Front RH = 10 mm/rear RH = 20 mm	2850
Front RH = 10 mm/rear RH = 25 mm	2880
Front RH = 15 mm/rear RH = 15 mm	2540
Front RH = 15 mm/rear RH = 20 mm	2580
Front RH = 15 mm/rear RH = 25 mm	2630
Front RH = 20 mm/rear RH = 15 mm	2400
Front RH = 20 mm/rear RH = 20 mm	2480
Front RH = 20 mm/rear RH = 25 mm	2540

in turns depends on the vertical loads, variable with front and rear RH, roll, yaw and steering angle, speed, longitudinal and lateral accelerations, etc.

As a further refinement, some aerodynamic transient loads are generated by variable car speed and variable front and rear RH that occurs when the car is maneuvering in violent braking, turn-in, and power-on phases.

Aerodynamic forces are measured in specialized wind tunnels equipped with rolling belts, able to simulate the ground traveling at the same speed as the air speed, or are computed by dedicated "computational fluid dynamics" (CFD) softwares [27,28]. It is also possible to measure the aerodynamic loads while the car is running, either at dedicated facilities instrumented with load cells embedded in the ground or directly at the race tracks with cars equipped with speed and acceleration sensors and with suspension load cells on the suspension elements.

Aerodynamic forces and their distribution between the front and the rear axle are also greatly influenced by traffic conditions, and for this reason, dedicated measurements are carried out with complex layouts in the wind tunnels or at the race tracks, to study the effects on cars running at a few lengths from each other, and within total or partial slip stream.

The big difficulty with outdoor testing is that the improvements sought are very small, in the order of a few percent between one complete car and its predecessor of the previous season, and the environment conditions are never fully controllable and repeatable. Therefore, outdoor testing, including coast-down tests [29], is becoming less and less relevant for racing applications because their cost is comparable with the cost associated with running a scale model in the wind tunnel or a virtual model with a CFD software.

14.4.3.2 Pitch Sensitivity

Another factor affecting the vehicle dynamics is the aerodynamic pitch sensitivity, generated by the variability of the aerodynamic forces with front and rear RH, and it is usually modeled by means of nonlinear skyhook speed-dependent springs and dampers that reduce the effective suspension stiffness and damping between the sprung mass and the unsprung masses [30].

Since the aerodynamic loads are speed dependent, the effect on the ride dynamics is the reduction of the static and the dynamic stability margins with speed, while the effect on handling is the redistribution of the vertical tire loads between the front and the rear axles [31].

14.4.3.3 Polar Curve and Aerodynamic Tuning

Downforce and drag forces can be adjusted by means of a great variety of parameters, like the angle of attack of the wing elements, gurney flaps, splitters, bargeboards, and inlet and exit radiator shutters; all these possible aerodynamic adjustments define an envelope curve that is referred as the car aerodynamic "polar" curve, since an excellent approximation of the curve is achieved with a second-order polynomial function relating total downforce to drag (Figure 14.4).

For a given set of car weight, engine power, and ambient conditions, the lap average speed for every race track is calculated for reasonable ranges of drag and downforce levels: Each pair of drag and downforce produces a defined result in terms of lap time, where more downforce usually increases the lap average speed up to the point that corners are coped easily at full throttle; conversely, more drag typically reduces the lap average speed as it reduces the net available longitudinal acceleration on the straights (Figure 14.5).

The lap time (or the lap average speed) is primarily a function of the car weight (full tank versus qualifying with minimum fuel) and of available engine power (race engine versus a qualifying engine, air density). The aerodynamic polar curve is a function of the ambient conditions through the air density factor. The fundamental task of the VDMR expert, when dealing with aerodynamic data, is to identify the optimal range of aerodynamic configurations that minimize the lap time for each race track, modeled as a peculiar set of cornering radii, straight-line lengths, local grip coefficient, banking, and slope changes.

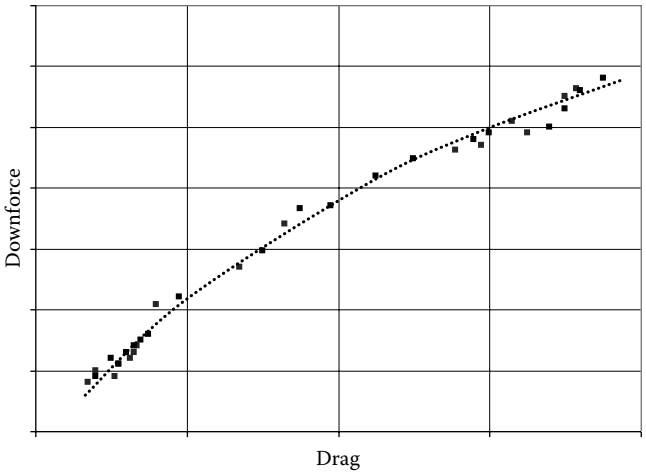


FIGURE 14.4 Polar curve as the envelope of all possible aerodynamic configurations and settings.

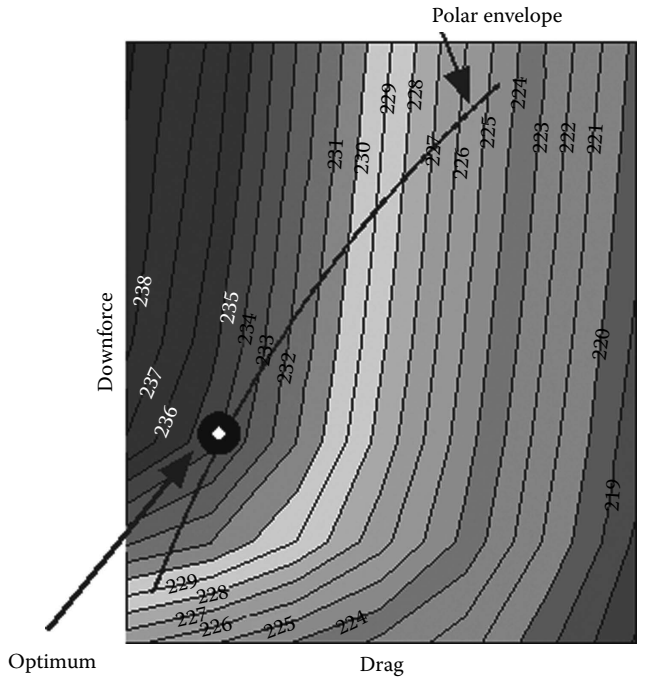


FIGURE 14.5 Polar curve and the average lap speed (km/h) matrix. Best aerodynamic configuration found.

At a more advanced level, the total downforce and front to rear aerodynamic balance have to be optimized to overcome a localized event on the race track, that is, to improve a local situation where the vehicle longitudinal or lateral or combined accelerations are limited by ride qualities over a local surface irregularity or by handling characteristics under severe maneuvers.

14.5 RACING CAR PERFORMANCE: LAP TIME

Lap time reduction is the ultimate goal of all people involved in racing [32].

Besides the aforementioned addressed engineering aspects, other hardly quantifiable aspects can affect lap time performance, especially for endurance races. They are the driver's

TABLE 14.7
Typical Lap Time Reduction (s)

	a	*b*	*c*
Weight decrease 1%	0.25	0.20	0.17
Center of gravity height decrease 1%	0.02	0.02	0.03
Level of grip (track/tire) increase 1%	0.30	0.32	0.35
Downforce increase 1%	0.05	0.07	0.08
Drag decrease 1%	0.13	0.12	0.10
Engine power increase 1%	0.15	0.14	0.12

psychology, the seat position, the visibility through the windscreen, the instinctive access to the car controls, and the cockpit temperature.

14.5.1 Car Performance

Let us base our lap time considerations on a typical 100 s lap (1 min 40 s).

For the same reference vehicles listed earlier (*a*, *b*, *c*), the following table shows the typical lap time reduction (in seconds) expected by improvements of the main performance parameters. Over a 100 s lap time, 1 s lap time reduction corresponds to one-hundredth performance improvement, and one-tenth of a second lap time reduction corresponds to one-thousandth performance improvement (Table 14.7).

From the aforementioned typical data, it appears that a one-hundredth improvement for the most important car parameters, like weight, engine power, tires, downforce, and drag, does not achieve anything close to a one-hundredth lap time improvement: all resulting lap time improvements are one order of magnitude smaller!

The aforementioned parameters are the *mean focus* of all vehicle dynamics analysis and do require very detailed and proven vehicle models as they drive the decision-making process of allocating the technical and human resources between design, structural analysis, wind tunnel testing, production, and track testing.

Also, each race track has distinctive lap time sensitivities to weight, engine power, tires, downforce, or drag adjustments.

These factors are generally constrained as it is difficult and very expensive to increase the engine power without losing reliability and it is not possible to run below the minimum legal weight or are interdependent as an increase in downforce is most of the time achieved at the expense of additional drag.

14.5.2 Car Development

14.5.2.1 Setup

"Setup" is a general term used in motor racing to define the set of all possible parameters/options/configurations of a racing car. Since the racing car is essentially an always evolving prototype, the setup has to be at the same time a rather flexible list in order to account for new development options and, on the other side, a very meticulous one in order to document, down to the small details, the particular configuration of the car so to reproduce its performance and baseline it to test further development components [2].

The setup has to be specific for the running conditions: the qualifying setup aims at letting the car achieve in a short-run maximum grip and ultimately the minimum single lap time even at the expense of high tire wear, partial driver discomfort, and temporary overstress on some car components.

The race setup is optimized to achieve the minimum overall race time, over a multiple sequence of long runs where it is difficult to predict the stabilized thermal working equilibrium of the four tires after ten or twenty consecutive laps.

In a very broad sense, the determining factor to achieve the minimum overall race time is the ability to predict, either by calculations or by experimental evidence, the performance decays of all systems, including the driver, over a prolonged time: typical and clear examples are the 24 h of Le Mans and the Indy 500.

Many factors do have an influence on lap time [2]; examples are given to highlight some conflicting requirements:

- Hotter ambient conditions do affect downforce and drag by the same amount through the decrease of the air density.
- Higher tire pressure reduces rolling resistance and wear but reduces grip.
- Less aero downforce increases the tire wear.
- Higher steering effort might allow the driver to get closer to the tire performance limit but can tire the driver.
- Higher camber angles might allow higher speed in the corner but increase wear and reduce straight-line performance in braking and acceleration.
- Less fuel load improves lap time due to the reduced overall weight but requires more frequent stops for refueling or the adoption of a leaner fuel injection engine map.

All vehicle dynamics aspects in a racing car, like aerodynamics, engine power, car weight, braking components, suspension components, and frame stiffness, ultimately rely on the tire properties to generate adequate grip, to wear evenly front to rear, and to work within a narrow temperature range.

The tire behavior ultimately dictates the final decisions over the desired handling properties, the driver's confidence, and the suspension and aerodynamics settings, especially in the long run (>30 laps).

The minimum lap time target depends also on the actual racing car trajectory ("the racing line"), which is hardly repeatable over the 10–14 corners of a typical road course circuit. To make the matter more complicated, the racing line is also influenced by the setup.

For example, more downforce allows a more violent braking into the corner and an earlier power application out of the corner and increases the cornering speed with effects on RH and on the suspension working point. The race line and the setup also depend on the particular driving style and on the characteristic skills of each driver.

Consequently, the minimum lap time might not have a unique solution or if unique, not a robust one [33].

14.5.2.2 Tires

The most important factor to optimize the lap time performance is the tire (Table 14.6). (For additional details on tires, see Part III.) The ability to predict how to tune the setup for the ambient conditions (level of grip of the track surface, track temperature, and ambient temperature) associated with the knowledge of the tire optimal working temperature, wear, and grip capability are the key factors to winning [34,35].

By taking the average time spent, over the course of one complete racing season, by single-seater cars racing on permanent circuits, it is possible to find out that for 45%–55% of the time, the car is in cornering situations, with high levels of yaw rate or yaw accelerations (steady-state or transient cornering maneuvers); for 35%–40% of the time, the engine is producing a fair amount of torque with significant level of longitudinal accelerations (on-throttle phases) and, finally, for 10%–15% of the time, the brakes are applied to produce violent longitudinal decelerations.

The total of the three aforementioned cases is more than 100% since the racing driver is always trying to use the engine power as early as possible out of the corner and to brake as late as possible into the corner.

While cornering, and in particular during transient cornering maneuvers, the front tires are stressed more than the rear tires; similarly, due to the heavy longitudinal load transfer while braking, the front tires are producing around 60% of the total braking forces.

It is therefore necessary, in order to optimize the tire load capacity, the grip demand, and the wear resistance between the front and the rear tires, to have the rear tires perform the driving function: this is one of the main reasons why the vast majority of racing cars are rear-wheel driven.

The tire is a combination of many elastomeric materials with a pronounced nonlinear ability to produce longitudinal and lateral grip based on the applied vertical load and on some geometry variables like camber, slip, and longitudinal slip. The geometry variables also influence the tire contact patch shape, tread pressure distribution, and wear. Moreover, the racing tire performance is dependent on the tread working temperature and on the micro- and macrotrack surface roughness, which is specific for every circuit.

Unfortunately, the generally available tire models do not describe the grip dependency from the temperature and from the surface roughness and therefore are of limited application in racing [36].

The role of the VDMR expert when dealing with tires is to build and tune the tire models to make the simulations representative of the reality, so as to produce reliable lap time predictions for the entire range of vehicle performance design and setup variables [34,37].

14.5.2.3 Driver

The racing driver is a very important factor in the overall performance. (For additional details on driver models, see Chapter 38.)

In order to achieve the highest performance of the vehicle/driver system, it is crucial to optimize the seat position, the easy access to all controls (gearshift, steering, throttle and brake pedals, switches, etc.), and the feedback from the steering wheel and the disturbances from the environment (front, rear, and side visibility, windscreen and helmet buffeting, etc.).

Two professional race drivers, driving the same chassis and the same engine for the same team, typically achieve a lap time performance within 0.2 s from each other over a 100 s lap time, so we can realize how efficient and repeatable machines the professional racing drivers are inasmuch as they are repeatable within 99.8%.

Racing, above all, is a mental sport.

The psychology and the character affect the driver performance over the course of a season and are the deciding factors for a professional racing driver to become a race winner or, ultimately, a champion. It is rather easy to win some races, but it is not easy to be consistently fast and bear the stress raised by the competitors and the technical and sporting challenges.

Every aspect on the car that reduces the level of stress of the driver while racing at the limit can be the subject of a vehicle dynamics study.

The professional racing driver is generally young, 8–14 years old in go-karting to 25–30 years old in single-seater Formulas like F3 to F1, and a vast majority of these professional racing drivers have had to stop, rather prematurely, attending the regular school classes in order to race.

The young age and the sometimes incomplete scholastic experience can make a driver reluctant to deeply learn some important technical aspects of the racing car and to stand criticisms, comments, and judgments or even to discriminate between true or false technical information.

Finally, in order to be successful, the professional racing driver has to have a very strong self-achievement motivation, sometimes even at the expense of his or her teammate or of all the people working in the same racing team.

Since the VDMR expert has to continuously deal with the interaction between the driver and the car, it is important for the VDMR expert to work together with the driver psychologist or the personal trainer, if available. In more general terms, the VDMR expert has to understand, properly value, and eventually influence the working relationships inside the racing team in order to let the racing driver become a mature, conscious, consistent, and reliable performer.

Steady-state situations like steering pads or maneuvers in prolonged constant radius corners do require a well-defined steering wheel angle to maintain the car in equilibrium at the maximum available speed and, therefore, do not depend much on the driver's ability. Non steady-state maneuvers, and racing is

all about handling in non steady-state maneuvers, require the driver to apply his or her controls (brakes, throttle, steering, gear shift), to judge on the car reactions, and sometimes to further finely adjust his or her inputs in a never-ending process [38]. Driver models are, therefore, very important to perform any type of race car simulations to the extent that it becomes difficult to discriminate between the car and the driver performance because the car and the driver are essentially a single symbiotic system [39,40].

Driving a race car, professionally and consistently, requires dedicated training and continuous race track practice, which are available at specialized racing schools [41–44].

The race driver has some characteristics in common with the aircraft pilot, like terminology, safety procedures, debrief, and even clothing and apparel [45], but there are some differences: A race car simulator is a luxury limited to the biggest Formula 1 teams and not available to the general driver; the racing driver has a rather short span and a generally limited technical preparation. Also, race driving is an individual sport, and there is no chance or reason why professional test drivers might develop the car for other racing drivers.

Racing is quite similar to a duel between multiple aircraft military pilots: the drivers and the pilots have to get ready for the race or the mission, properly determined in their mind to overcome at any cost the opponents.

Racing drivers and military aircraft pilots have to conserve their rationale up to the point that to win a race or to survive a military flight duel, they cannot hate their competitor or enemy as this irrational feeling absorbs mental energy and makes them focus too much on the details and finally loses their spatial perception of the global environment and the time-ordered hierarchy of their actions and reactions.

To be a successful race driver or an excellent military aircraft pilot, it is necessary to psychologically stress the competitors, to make them lose their concentration, and finally, to make a mistake.

It is very instructive, at the end of a race, to watch the drivers leave their cars: those drivers who are able to show a calm attitude, to cheer to the fans, and to their competitors have greater chances to win than nervous drivers who lost their temper because of a racing accident or simply because everything did not go as desired.

14.5.2.4 Driving Simulator for Race and High Performance Cars

Driving simulators are increasingly used for race applications [47]. They can be particularly useful for driver coaching. They can also be used for the development of setups (e.g., to assess tire thermal effects for both qualifying and race setups) and for data analysis (e.g., to analyze tire forces).

A new generation driving simulator (Figure 14.6) features nine degrees of freedom and is designed to take full advantage of the motion cueing algorithm [48–51] based on an MPC approach



FIGURE 14.6 Driving simulator at Ferrari.

with in-the-loop real-time optimization. It consists of a hexapod mounted on a triangular-shaped frame, sliding on an even and stiff steel surface by means of a combination of air and magnetic bearings. The hexapod generates vertical, pitch, and roll movements, as well as small yawing movements and movements in longitudinal and lateral directions. The sliding frame makes the driver feel the longitudinal, lateral, and yaw accelerations. The driving simulator platform is suitable for testing both vehicle dynamics and ride comfort.

14.6 CONCLUSIONS

VDMR is an engineering discipline that covers many aspects of the racing car performance and therefore has a strategic role, from the initial concept down to the final tuning, to help the designer, the production manager, the race engineer, and the aerodynamicist optimize the overall lap time performance and to ensure adequate handling and ride qualities.

The racing driver is a fundamental part of the driver/vehicle system, and his or her individual driving preferences are pivotal to the overall lap time. For these reasons, vehicle dynamics has to translate the subjective feedback from the racing driver into objective performance indexes.

REFERENCES

1. Fey, B., *Data Power—Using Racecar Data Acquisition*, Robert Towery, Memphis, TN, 1993.
2. Milliken, W., *Race Car Vehicle Dynamics*, SAE, Troy, MI, 1995.
3. Trevorrow, N., Kuhnell, B., and Preston, M., *Track Surface Effects on Race Tyre Performance*, SAE, Troy, MI, 2004, Paper 2004-01-3527.
4. Tokutake, H., Miura, Y., and Okubo, H., *Workload Analysis Method via Optimal Driver Model*, SAE, Troy, MI, 2004, Paper 2004-01-3536.
5. Wright, P., *Ferrari Formula 1—Under the Skin of the Championship Winning F1-2000*, SAE, Troy, MI, 2004.
6. FIA website like for Motorsport Regulations, www.fia.com/sport/Regulations/index.html. (Accessed July 2013)
7. Milliken, F., Whitcomb, D., Segel, L. et al., *Research in Automobile Stability and Control and in Tyre Performance*, Institution of Mechanical Engineers/Cornell Aeronautical Laboratory, Buffalo, NY, 1956.
8. Etkin, B., *Dynamics of Flight*, John Wiley & Sons, New York, 1959.
9. Irving, F.G., *An Introduction to the Longitudinal Static Stability of Low Speed Aircraft*, Pergamon Press, Oxford, U.K., 1966.
10. Metz, L.D., *What Constitutes Good Handling*, SAE, Troy, MI, 2004, Paper 2004-01-3532.
11. Ellis, J.R., *Road Vehicle Dynamics*, published by the author, 1989.
12. Mitchell, W., Schroer, R., Unger, M., and Grisez, D., *Training Test Drivers with Data Acquisition*, SAE, Troy, MI, 2000, Paper 2000-01-3568.
13. Rice, R.S., *Measuring Car-Driver Interaction with the G-G Diagram*, SAE, Troy, MI, 1973, Paper 730018.
14. Mitchell, W.C., Schroer, R., and Grisez, D.B., *Driving the Traction Circle*, SAE, Troy, MI, 2004, Paper 2004-01-3545.
15. Gillespie, T., *Fundamentals of Vehicle Dynamics*, SAE, Troy, MI, 1992.
16. Morse, P., *Using K&C Measurements for Practical Suspension Tuning and Development*, SAE, Troy, MI, 2004, Paper 2004-01-3547.
17. Gobbi, M. and Mastinu, G., Analytical description and optimization of the dynamic behavior of passively suspended road vehicles, *Journal of Sound and Vibrations*, 245(3), 457–481, 2001.
18. Kelly, J., Kowalczyk, H., and Oral, H.A., *Track Simulation and Vehicle Characterization with 7 Post Testing*, SAE, Troy, MI, 2002, Paper 2002-01-3307.
19. Thompson, L.L., Soni, H., and Law, H., *The Effect of Chassis Flexibility on Roll Stiffness of a Winston Cup Race Car*, SAE, Troy, MI, 1998, Paper 983051.
20. Thompson, L., Raju, S., and Law, H., *Design of a Winston Cup Chassis for Torsional Stiffness*, SAE, Troy, MI, 1998, Paper 983053.
21. Smith, C., *Racing Chassis and Suspension Design*, SAE, Troy, MI, 2004.
22. Smith, C., *Tune to Win*, Aero Publishers, Inc., Fallbrook, CA, 1978.

23. Segel, L., *Theoretical Prediction and Experimental Substantiation of the Response of the Automobile to Steering Control*, Institute of Mechanical Engineers, London, U.K., 1956.
24. Abbott, I.H. and Von Doenhoff, A.E., *Theory of Wing Sections*, McGraw-Hill Book Co, New York, 1949.
25. Hucho, W.H., *Aerodynamics of Road Vehicles from Fluid Dynamics to Vehicle Engineering*, SAE, Troy, MI, 1998.
26. Hartley, J., *Tyre Characteristics as Applicable to Vehicle Stability Problems*, Institute of Mechanical Engineers, London, U.K., 1953–1954.
27. Carr, G., *A Comparison of the Ground Plane Suction and Moving Belt Ground Representation Techniques*, SAE, Troy, MI, 1988, Paper 880249.
28. Pope, A. and Harper, J.J., *Low Speed Wind Tunnel Testing*, John Wiley & Sons, New York, 1966.
29. Buckley, F., *An Improved Coast Down Test and Method Analysis*, SAE, Troy, MI, 1995, Paper 950626.
30. Doniselli, C., Mastinu, G., and Gobbi, M., *Aerodynamics Effects on Ride Comfort and Road Holding of Automobiles—Vehicle System Dynamics Supplement 25*, Swets & Zeitlinger, Amsterdam, the Netherlands, 1996.
31. Haney, P. and Braun, J., *Inside Racing Technology*, TV Motorsport, Redwood City, CA, 1995.
32. Deakin, A. and Crolla, D., *Fundamental Parameters Design Issues which Determine Race Car Performance*, SAE, Troy, MI, 2000, Paper 2000-01-3537.
33. Hsieh, C., *Automotive Applications Using a Robust Design Approach*, SAE, Troy, MI, 1999, Paper 1999-01-1029.
34. Haney, P., *The Racing & High Performance Tyre, Using the Tyres to Tune for Grip and Balance*, SAE, Troy, MI, 2003.
35. Van Eldik, T. and Pacejka, H.B., *The Tyre as a Vehicle Component*, Delft University of Technology, Delft, the Netherlands, 1971.
36. Pacejka, H.B., *Tyre Characteristics and Vehicle Dynamics*, Delft University of Technology, Delft, the Netherlands, 1986.
37. Fonda, A.G., *Tyre Tests and Interpretation of Experimental Data*, Institute of Mechanical Engineers, London, U.K., 1956.
38. Siegler, P., Crolla, D., and Deakin, A., *Lap Time Simulation: Comparison of Steady State, Quasi Static, and Transient Racing Car Cornering Strategies*, SAE, Troy, MI, 2000, Paper 2000-01-3563.
39. Kroll, C.V. and Roland, R.D. Jr., A preview-predictor model of driver behavior in emergency situations, Cornell Aeronautical Laboratory Report VJ-2251-V-6, Buffalo, NY, 1970.
40. Day, T. and Metz, D., The simulation of driver inputs using a vehicle driver model, SAE, Troy, MI,2000. Paper 2000-01-1313.
41. Lopez, C., *Going Faster! Mastering the Art of Race Driving: The Skip Barber Racing School*, Robert Bentley, Cambridge, MA, 1997.
42. Senna, A., *Principle of Race Driving*, Hazleton Publishing, London, U.K., 1991.
43. Donohue, M., *Van Valkenburg P: The Unfair Advantage*, Robert Bentley, Cambridge, MA, 1975 (revised edition 2000).
44. Moss, S. and Pomeroy, L., *Design and Behavior of the Racing Car*, William Kimber, London, U.K., 1963.
45. Cooper, G. and Harper, R., The use of pilot rating in the evaluation of aircraft handling qualities, NASA report TN D-5153, 1970.
46. http://www.dallara.it/simulator.php
47. Telban, R.J., Wu, W., Cardullo, F.M., and Langley Research Center. *Motion Cueing Algorithm Development: Initial Investigation and Redesign of the Algorithms*, National Aeronautics and Space Administration, Langley Research Center, Hampton, VA, 2000.
48. Baseggio, M., Beghi, A., Bruschetta, M., Maran, F., and Minen, D., An mpc approach to the design of motion cueing algorithms for driving simulators, in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pp. 692–697, October 2011.
49. Beghi, A., Bruschetta, M., Maran, F., Minen, D., A model-based motion Cueing strategy for compact driving simulation platforms, *Driving Simulation Conference*, Paris, France, 2012.
50. Maran, F., Model-based control techniques for automotive applications, PhD Thesis, University of Padua, Padua, Italy, 2013.

15 Race Cars
Braking System

Carlo Maria Domenico Cantoni and Giorgio Previati

CONTENTS

15.1 Introduction 517
15.2 F1 Car Behavior under Straight-Line Braking 517
15.3 F1 Car Braking System Layout 521
15.4 Brake Energy Dissipation during a Lap 525
15.5 Brake Caliper Design and Friction Material 526
15.5.1 Caliper Design 526
15.5.2 Friction Material 527
15.6 Conclusions 528
References 528

15.1 INTRODUCTION

The main purpose of this chapter is to explain the importance and complexity of the brake system for a racing car.

The importance of the braking system refers to both reducing the number and severity of accidents and reducing lap times. Hence, braking is one of the most critical areas of racing car.

In this chapter, the experience and data gathered by some of the authors in Formula 1 (F1) races are included. It is important to highlight that the numbers reported in the following tables are representative of an F1 car of years 1999–2002, but the approach and the modeling remain relevant today.

This chapter is divided into four main topics:

1. F1 car behavior under straight-line braking
2. F1 braking system layout
3. Brake energy dissipation during a single lap
4. Brake caliper design and friction material characteristics

Braking is important for race cars because shorter stopping distances also mean that the engine can be at full power for a longer time and the car's lap time can be dramatically improved. (For more details on race cars, see Chapter 14.)

15.2 F1 CAR BEHAVIOR UNDER STRAIGHT-LINE BRAKING

The main phenomena that affect the car during the braking maneuver are addressed.

The mathematical formula that gives the ideal braking force that the car can exert and the ideal braking balance (front-to-rear ratio) during braking is given.

In Table 15.1, the main parameters for the description of a braking F1 car are reported along with their respective typical values.

TABLE 15.1
Car Parameters

Parameter	Typical Value	Name
μ	1.5	Tire friction, constant and equal for all tires
m	600 (kg)	Mass (including driver, empty condition)
$\alpha = \frac{Z_f}{Z_f + Z_r}$	0.42	Front weight distribution
h	0.285 (mm)	Center of gravity height
h_{cp}	0.285 (mm)	Center of pressure height
l	2955 (mm)	Wheelbase
ρ	1.225 (kg/m^3)	Air density
S	1.45 (m^2)	Car frontal area
C_{zT}	2.4	Total lift coefficient
$k_a(v) = k_1 v^2 + k_2 v + k_3$	$k_1 = 0$ (s^2/m^2) $k_2 = 0$ (s/m) $k_3 = 0.35$	Aerodynamic balance of the car as ratio between front and total downforce. In general, it is function of speed (v).
$C_{zf} = k_a \times C_{zT}$	0.84	Front lift coefficient
$C_{zr} = (1 - k_a) \times C_{zT}$	1.56	Rear lift coefficient
C_x	0.98	Drag coefficient

Source: B. Breuer and K.H. Bill, *Brake Technology Handbook*, SAE, Troy, MI, 2008.

During a braking maneuver, three important aerodynamic forces are acting on the car. These forces are also present in a normal production car but due to their very low values are normally neglected. These three forces (aerodynamic downforces at the front and rear axles, respectively, and aerodynamic drag at the center of gravity) affect the overall braking performance of the car.

Referring to Figure 15.1 and considering the parameters of Table 15.1, the following formulas describe longitudinal (Equation 15.1) and vertical (Equation 15.2) equilibrium and rotational equilibrium around front (Equation 15.3) and rear (Equation 15.4) contact points:

$$X_f + X_r + F_d = ma \tag{15.1}$$

$$Z_f + Z_r = F_{lf} + F_{lr} + mg \tag{15.2}$$

$$Z_f l - F_{lf} l - mgl_r - mah + F_d h_{cp} = 0 \tag{15.3}$$

$$Z_r l - F_{lr} l - mgl_f + mah - F_d h_{cp} = 0 \tag{15.4}$$

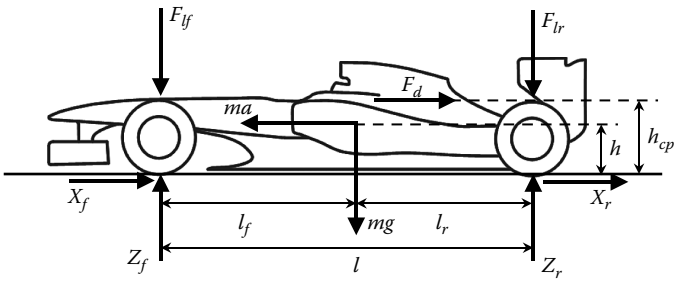


FIGURE 15.1 Forces acting on a racing car during braking.

When approaching locking conditions, $X_f = \mu Z_f$ and $X_r = \mu Z_r$. From Equations 15.3 and 15.4, the longitudinal forces can be computed as

$$X_f = \mu Z_f = \mu m\left(g\frac{l_r}{l} + a\frac{h}{l}\right) + \mu F_{lf} - \mu F_d\frac{h_{cp}}{l} \tag{15.5}$$

$$X_r = \mu Z_r = \mu m\left(g\frac{l_f}{l} - a\frac{h}{l}\right) + \mu F_{lr} + \mu F_d\frac{h_{cp}}{l} \tag{15.6}$$

From Equations 15.1 and 15.2, the longitudinal acceleration can be computed as

$$X_f + X_r = \mu(Z_f + Z_r) = ma - F_d \Rightarrow a = \frac{\mu}{m}(Z_f + Z_r) + \frac{F_d}{m} = \frac{\mu}{m}(F_{lf} + F_{lr} + mg) + \frac{F_d}{m} \tag{15.7}$$

Finally, by replacing Equation 15.7 in Equations 15.5 and 15.6, the front and rear longitudinal forces read

$$X_f = \frac{\mu mg}{l}(l_r + \mu h) + \mu\left(\mu\frac{h}{l}(F_{lf} + F_{lr}) + F_{lf} - F_d\frac{h_{cp} - h}{l}\right) \tag{15.8}$$

$$X_r = \frac{\mu mg}{l}(l_f - \mu h) + \mu\left(\mu\frac{h}{l}(F_{lf} + F_{lr}) + F_{lr} + F_d\frac{h_{cp} - h}{l}\right) \tag{15.9}$$

Equations 15.8 and 15.9 can be rewritten as a function of the car parameters by considering the expression of the aerodynamic lift forces* and the weight balance of the car:

$$X_f = \mu mg\left(\alpha + \mu\frac{h}{l}\right) + \frac{1}{2}\rho v^2 S\mu\left(\mu\frac{h}{l}C_{zT} + C_{zF} - C_x\frac{h_{cp} - h}{l}\right) \tag{15.10}$$

$$X_r = \mu mg\left(1 - \alpha - \mu\frac{h}{l}\right) + \frac{1}{2}\rho v^2 S\mu\left(-\mu\frac{h}{l}C_{zT} + C_{zR} + C_x\frac{h_{cp} - h}{l}\right) \tag{15.11}$$

The brake balance of the car can be defined as

$$F_{B,dist}(v,\mu) = \frac{X_f(v,\mu)}{X_f(v,\mu) + X_r(v,\mu)} \tag{15.12}$$

Equations 15.10 through 15.12 give the ideal brake force distribution, provided that the friction is equal at front and rear tires and that front and rear tires work at adherence limit.

The considered equations have been derived from the equilibrium of the vehicle; thus, this formulation is still valid in case car parameters vary during braking, provided that the constant parameters are replaced by the corresponding functions. In this case, actual tire properties (friction variation due

* The aerodynamic lift forces can be computed as [3]

$$F_{lf} = \frac{1}{2}\rho C_{zF} S v^2, \quad F_{lr} = \frac{1}{2}\rho C_{zR} S v^2, \quad F_{lT} = F_{lf} + F_{lr} = \frac{1}{2}\rho C_{zT} S v^2, \quad F_D = \frac{1}{2}\rho C_x S v^2$$

to temperature, vertical load, and rolling speed), realistic and speed-dependent aerodynamic data, car suspension geometry (antidive effects), and center of gravity and wheelbase variations due to load variation have to be taken into account. Very complex models have to be developed in order to derive the functions describing the dependence of the car parameters on such variations.

Referring to aerodynamic forces (lift and drag coefficients), it has been proven, both mathematically and experimentally [1,2], that such forces are functions of the attitude of the car. In particular, during braking, a certain amount of pitch and heave is present, depending on suspension and tire stiffness, varying the height of the front and rear of the vehicle. In [3], it has been shown that variation of the front and rear heights of the vehicle (depending on the heave of the center of gravity and on the pitch angle) are responsible of sensible variations in the front and rear lift coefficients. The maps of the kind of the ones shown in Figure 15.2 can be used to relate the lift coefficients to the current front and rear heights during the maneuver. Experience has shown that in case of a braking maneuver of an F1 car, the variation of these coefficients has effects on the dynamic of the vehicle up to the second derivative.

In Figure 15.3, the ideal braking balance of an F1 car has been depicted by considering the variations of friction and aerodynamic coefficients during a braking maneuver from 300 to 120 km/h.

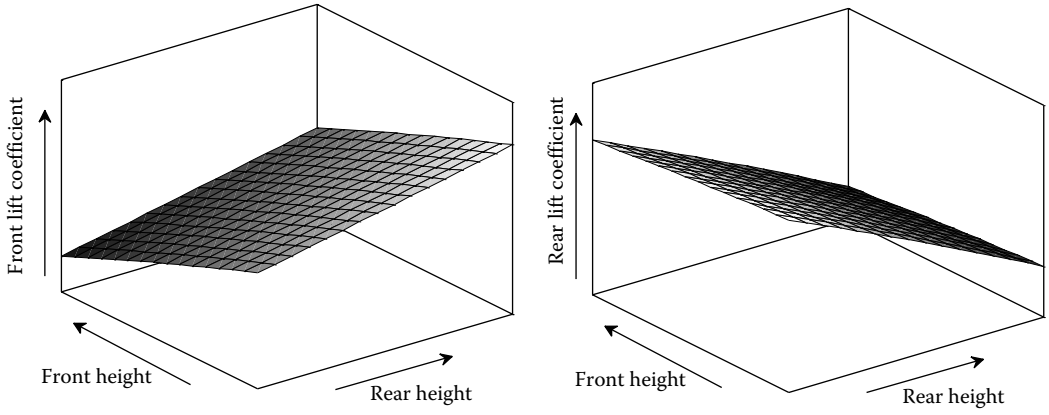


FIGURE 15.2 Maps for front and rear lift coefficients as a function of front and rear height. (From C. Doniselli et al., Aerodynamic effects on ride comfort and road holding of automobiles, *Proceedings of the XIV IAVSD Symposium*, Swets & Zeitlinger, Ann Arbor, MI, 1995.)

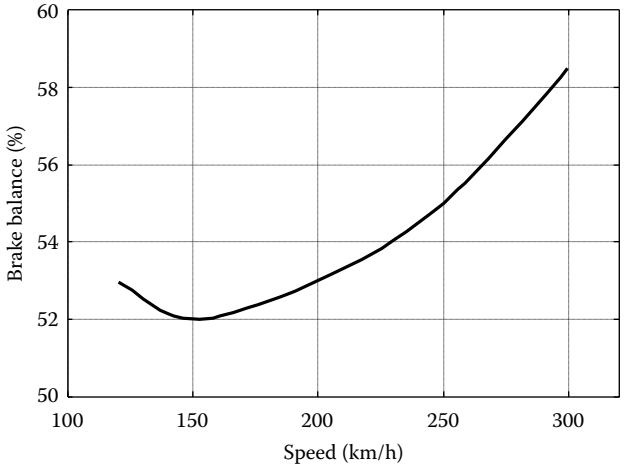


FIGURE 15.3 Ideal brake balance distribution for an F1 car.

In comprehensive models, some other extra braking forces acting on the car have to be considered. These forces, mostly in the direction of the aerodynamic drag, act on the rear axle and are due to the engine that provides substantial braking torque during the braking maneuver, being the throttle not open in this situation. This enhances the braking at the rear axle by means of the torque transmitted by the gearbox and the final drive ratio.

As depicted in Figure 15.3, the theoretical braking force balance between front and rear of an F1 car is not constant. That is, the ratio between front and rear braking forces varies during the braking maneuver.

At higher speed, more front braking force is required for car stability. At lower speed, the brake balance is shifted toward the rear axle to avoid front wheel locking.

The general tendency of front-to-rear brake balance and its speed dependency has been confirmed by practical track testing during the last years.

15.3 F1 CAR BRAKING SYSTEM LAYOUT

Figure 15.4 shows the typical layout of a braking system for an F1 car. The two master cylinders are connected via a kinematic device, namely, the balance bar, which distributes the pedal force to the pushrods of the two master cylinders and thus to the two separate circuits for the front and rear brakes. The pedal works as a lever, increasing the force applied by the driver, with a ratio between 3:1 and 4:1. The choice of master cylinder diameters influences the driver's effort and his or her feedback in terms of the force applied and pedal travel. In most cases, the brake master cylinders are fed by two separate reservoirs.

The hydraulic circuit is realized by means of two lines starting from each master cylinder and specific to each axle. The lines consist of a partly rigid and a partly flexible piping with diameters from 2 to 3 mm. The two circuits terminate at the brake calipers.

The caliper clearly plays a fundamental role in the brake system, and it has several requirements. Firstly, the caliper has to be light, as it adds to the weight of unsprung mass. A stiff caliper is essential for assuring that the action of the fluid under pressure is used for generating braking power, rather than deforming the caliper itself. It should also protect the brake fluid from high temperatures caused by the friction between the pads and the disk. Finally, it must guarantee a high level of safety by maintaining easy bleeding.

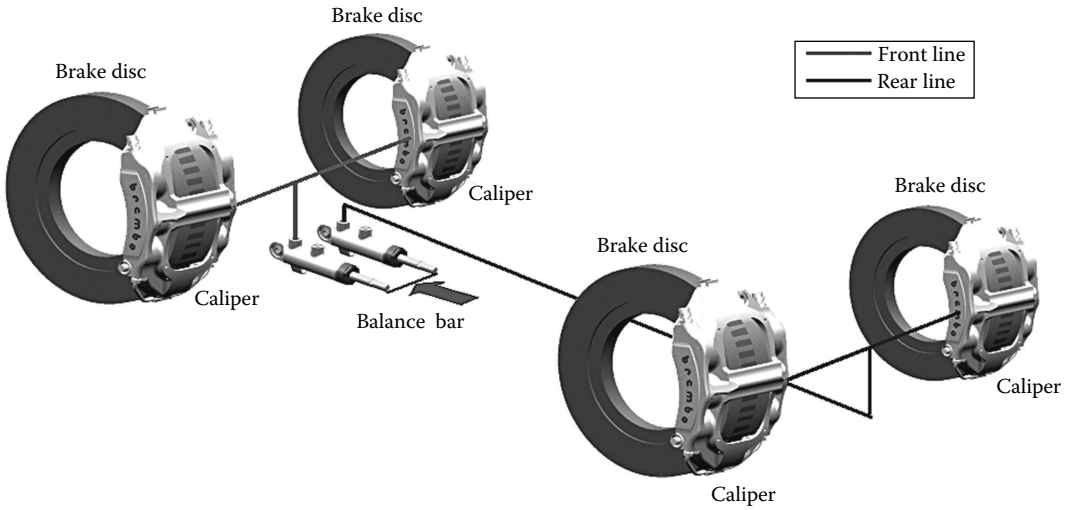


FIGURE 15.4 Typical brake system configuration. (Adapted from B. Breuer and K. H. Bill, *Brake Technology Handbook*, SAE, Troy, MI, 2008.)

The disk of racing cars is usually realized in two parts, the friction surface and the bell. The friction surface is the part devoted to the contact with the pads, and the bell represents the drive element of the disk.

In the case of an F1 car, as it can be noticed from Figure 15.3, no ABS hydraulic unit is present, as the current Federation Internationale de l'Automobile (FIA) regulations do not allow any form of electronically controlled systems in the braking area. In the past, in early 1990s, some ABS units were fitted in F1 cars with very encouraging results, but subsequent FIA regulations stopped this research field in order of cost reductions. Currently, no form of driver aids is allowed in the braking system, the braking performance of the car must be under the exclusive control of the driver, and even servo units are not allowed and thus not present in the system layout. The result is a very simple braking system. The steps for the generation of braking torque are the following:

- The driver supplies mechanical energy to the system by pushing the pedal. The energy supplied by the driver (E_D) can be computed as

$$E_D = F_P T_p$$

  where
  F_p is the force applied by the driver
  T_p is the travel of the pedal

- The pedal ratio multiplies the force exerted by the driver
- Brake pressure is generated inside each of the master cylinders, according to the displacement of the corresponding rod
- The pressure is transmitted to the brake caliper where the brake torque is generated by means of the friction between brake disk and pads

The brake balance of the car is obtained by means of the balance bar placed between pedal and master cylinders. The driver can change the brake balance by moving the balance bar with a mechanical rotary switch. The balance bar acts as a lever, and a change of its position changes the lever ratio between the two rods giving a different ratio between front and rear pressure. In case the balance bar is in its middle position and the two master cylinders have the same size, the same level of pressure is obtained in the two circuits. If the balance is placed closer to one of the master cylinders, in the correspondent circuit, a higher brake pressure is obtained than in the other one.

By remembering Equations 15.10 through 15.12, ideal pressures for front and rear braking circuits can be computed as follows:

$$P_f(v,\mu) = \frac{X_f(v,\mu) r_f}{2 r_{df} \mu_{disc} A_f} \tag{15.13}$$

$$P_r(v,\mu) = \frac{X_r(v,\mu) r_r}{2 r_{dr} \mu_{disc} A_r} \tag{15.14}$$

$$P_{tot}(v,\mu) = P_f(v,\mu) + P_r(v,\mu) \tag{15.15}$$

$$P_{dist}(v,\mu) = \frac{P_f(v,\mu)}{P_f(v,\mu) + P_r(v,\mu)} \tag{15.16}$$

The total pressure (P_{tot}) defined in Equation 15.15 does not have any physical meaning. However, since the two master cylinders have the same (or very similar) area, this quantity is proportional to

TABLE 15.2
Brake System Parameters

Parameter	Typical Value	Name
A_f	5800 (mm^2)	Front caliper area
A_r	5000 (mm^2)	Rear caliper area
r_{df}	117 (mm)	Front disk effective radius
r_{dr}	117 (mm)	Rear disk effective radius
r_f	315 (mm)	Front tire effective radius
r_r	317 (mm)	Rear tire effective radius
μ_{disk}	0.45	Friction coefficient between disk and pad

the pedal force, and it can be used as an index of the utilization of the brake system. The explanation of the symbols employed in the equation is reported in Table 15.2 along with their typical values. Equations 15.13 through 15.16 are very informative regarding

- Ideal front brake pressure for ideal front braking force (Equation 15.13)
- Ideal rear brake pressure for ideal rear braking force (Equation 15.14)
- Ideal total pressure for total braking force for the car. From the value of the total pressure, the total energy and power can be easily computed
- Ideal brake pressure balance for ideal car brake distribution

Figure 15.5 shows the ideal car brake balance and the ideal pressure balance as a function of the total braking pressure. These pictures can be obtained by combining the equations for the computation of the braking forces (Equations 15.10 through 15.12) and the equations for the computations of the brake pressures (Equations 15.13 through 15.16). Higher total pressure corresponds to higher car speed.

In particular, the knowledge of Figure 15.5b is most important for the design of the braking system of the car. In fact, the brake system should be designed in order to achieve a brake distribution (i.e., front and rear pressure and brake pressure balance) as close as possible to the one shown in this figure.

In the case of F1 cars, under current FIA rules and constraints, only a constant pressure balance can be employed. In Figure 15.6, the comparison between a fixed pressure balance (horizontal line) and the ideal pressure balance is shown. By inspecting this figure, one can notice that the choice of the fixed level of pressure balance is a compromise between the different needs at higher and lower speeds. By adjusting the balance bar position, the driver has the opportunity to find the best compromise for each braking maneuver according to current track conditions and tire wear.

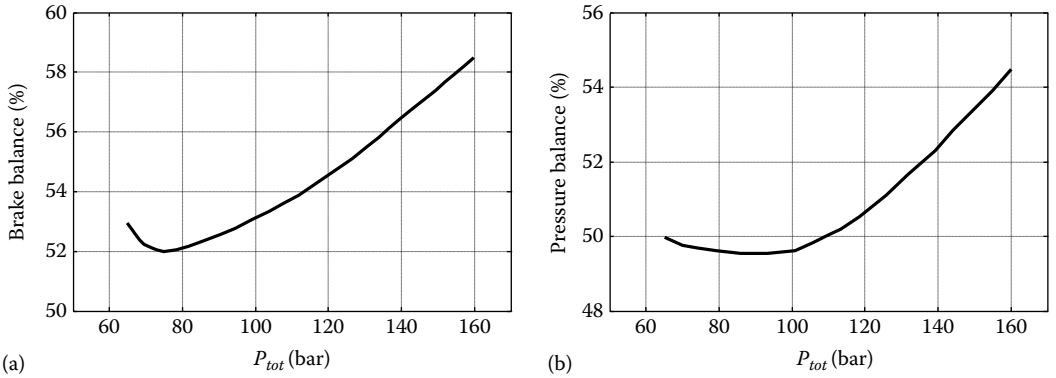


FIGURE 15.5 (a) Ideal brake balance as a function of the total braking pressure. (b) Ideal pressure balance as a function of the total braking pressure.

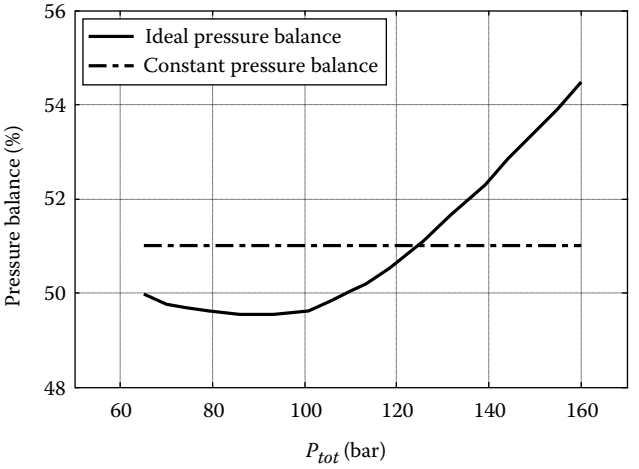


FIGURE 15.6 Ideal braking balance as a function of the total braking pressure and comparison with a constant balance situation.

From a technological point of view, a brake distribution very close to the ideal one could be obtained by employing electronic valves and controls. Currently, this is not allowed in F1 races, as FIA has decided to leave the braking system under the sole control of the driver.

Considering a fixed brake balance β, the effective rear brake force can be computed as (see Equation 15.12)

$$X_{rb}(v,\mu) = \frac{X_f(v,\mu)}{\beta} - X_f(v,\mu) \tag{15.17}$$

The car deceleration is

$$a(v,\mu) = \frac{X_f + X_{rb} + F_d}{m} = \frac{X_f \beta^{-1} + F_d}{m} \tag{15.18}$$

In [4], it is pointed out that this simplified formulation can give very good approximation when compared with experimental data. In Figure 15.7, the predicted acceleration (Equation 15.18) is compared with the acceleration measured by telemetry during an actual braking.

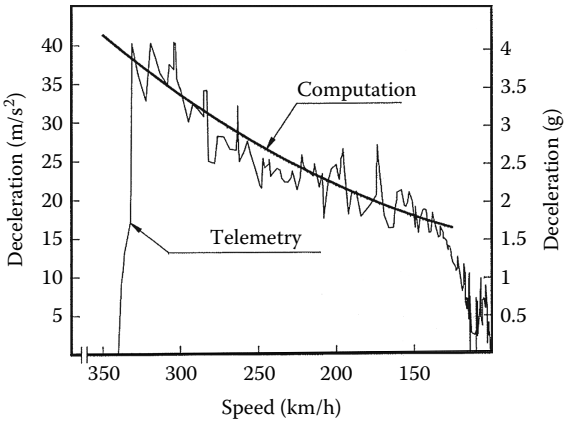


FIGURE 15.7 Comparison between measured and computed deceleration of an F1 car during braking. (Adapted from E. Benzing, *Dall'aerodinamica alla Potenza in Formula 1*, Giorgio Nada Editore, Vimodrone, Italy, 2004.)

15.4 BRAKE ENERGY DISSIPATION DURING A LAP

In this section, the circuit of Melbourne, current location of the F1 Grand Prix of Australia, is considered as reference track for the computation of the energy dissipated by the braking system for a single lap and for the computation of the required braking power. Figure 15.8 shows a map of the reference track on which the most significant speeds, corners numbering, and sectors times are reported.

For a standard vehicle (i.e., no aerodynamic surfaces) on a plane road, the kinetic energy is completely dissipated by the braking system, being negligible the energy dissipated by the aerodynamic drag. In this case, the dissipated energy can be computed as the difference between the kinetic energy before and after braking.

In case of a racing car, the energy dissipated by the aerodynamic drag must be considered. For the computation of the energy dissipated by the braking system, the deceleration of the vehicle (Equation 15.18), depending on speed, has to be integrated between initial and final velocities. For each integration instant, braking forces (Equations 15.10 and 15.17) and aerodynamic drag have to be computed. Brake and drag instantaneous power can be estimated by multiplying these forces by the current velocity. Finally, brake and drag dissipated energy can be obtained by time integration of the previously computed powers. Figure 15.9 shows brake and drag dissipated energy and power computed for a braking maneuver from 300 to 100 km/h by considering all the parameters of the car as constants and by neglecting the dynamic of the brake system (i.e., the brakes apply instantaneously all the required braking torque to have the maximum braking forces). The last hypothesis overestimates the braking power in the first part of the maneuver. The brake balance has been considered fixed and equal to 0.53.

In Table 15.3, computed values of brake dissipated energy are reported for the significant braking maneuvers of the reference track, computed by a complete model considering all the main effects of the braking system. In the considered case, eight relevant braking actions can be identified at the corners reported in the table.

By inspecting the values of Table 15.3, one can immediately have an idea of the huge amount of energy that the brake system has to dissipate in a very short time.

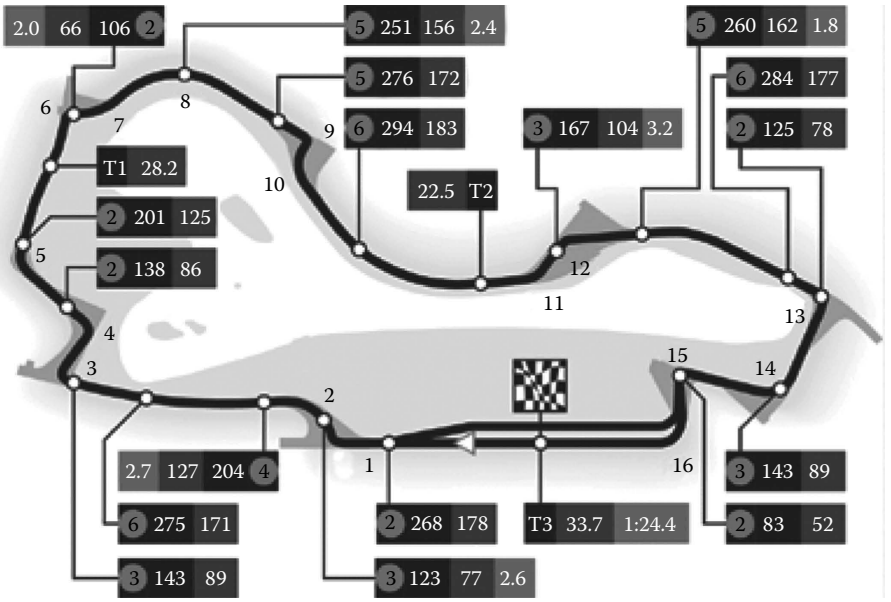


FIGURE 15.8 Melbourne circuit (reference track).

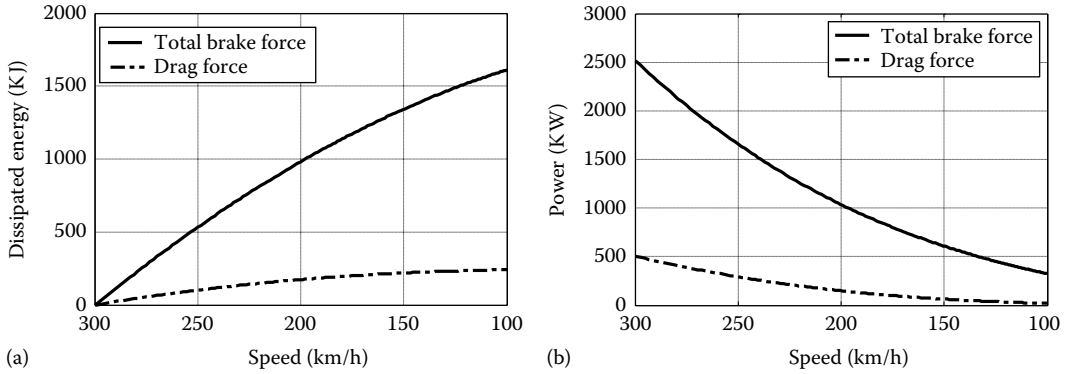


FIGURE 15.9 (a) Computed brake and drag dissipated energy. (b) Computed brake and drag power. Data in Table 15.1.

TABLE 15.3
Brake Dissipated Energy and Braking Power for the Reference Track

Corner Number (No. 1, 2, …, 15)	Initial Speed (km/h)	Final Speed (km/h)	Braking Time (s)	Braking Energy (kJ)	Braking Power (Peak Value) (kW)
1	320	156	1.60	1.30×10^3	1.66×10^3
3	308	93	2.35	1.43×10^3	1.58×10^3
6	301	146	1.50	1.10×10^3	1.34×10^3
9	300	114	1.90	1.28×10^3	1.62×10^3
12	312	250	0.80	0.52×10^3	1.02×10^3
13	313	160	1.50	1.22×10^3	1.66×10^3
14	245	213	0.60	0.21×10^3	0.57×10^3
15	254	106	1.75	0.87×10^3	0.99×10^3

Corners numbering in Figure 15.8.

In particular situations, severe circuit or high grip configurations, the braking power can exceed 2000 horse power. For withstanding such huge powers and energy dissipation, which generates high thermal and mechanical stresses, the brake system design must be extremely accurate and robust, especially at wheel level.

15.5 BRAKE CALIPER DESIGN AND FRICTION MATERIAL

15.5.1 Caliper Design

In designing racing brake calipers, two fundamental requirements have to be considered: stiffness, crucial for obtaining short pedal strokes, and lightness, important in racing cars where mass distribution strongly influences performances.

As a result, brake calipers for race cars are usually very stiff. In this situation, floating disk and bell assemblies are employed to provide the necessary clearance.

The design of the brake caliper, in terms of stiffness and lightness, depends strongly on the rules of the championship they are designed for. In most cases, rules impose limits on brake systems, in terms of allowed structures and materials. Referring to the F1 championship, the material used for manufacturing brake calipers must have an elastic modulus of less than 80 GPa; this requires alloys with a stiffness as close as possible to this limit but with the lowest possible density.

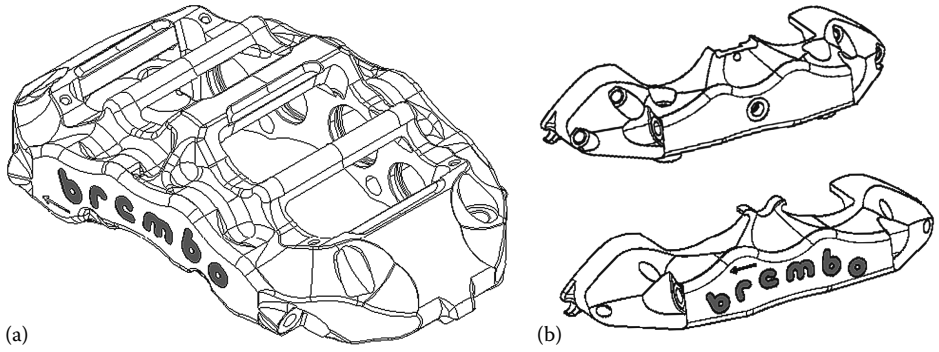


FIGURE 15.10 Brake calipers. (a) Monobloc caliper. (b) Two-piece caliper. (Adapted from B. Breuer and K. H. Bill, *Brake Technology Handbook*, SAE, Troy, MI, 2008.)

A caliper is composed of

Caliper body. A monobloc is usually employed, derived from a billet or a forging. Alternatively, a two-piece assembly can be used, which is usually machined from an aluminum billet (Figure 15.10). The caliper undergoes hard anodizing or nickel-plating or other surface treatments in order to make it resistant to possible collisions (stones, rubble, etc.).

Pistons are made from aluminum and subjected to anodizing. A titanium insert is used as isolation from heat coming from the pad. The machining tolerances are very tight.

Crossover pipe. Usually made of stainless steel, it allows the fluid to reach the caliper side opposite the inlet side [5].

Seals are very important for controlling roll-back, i.e., piston dynamics, and for guaranteeing a hydraulic seal even at high temperatures. If the seal does not allow the piston to return to its resting position, some residual torques arise between the pads and disk. Vice versa, if the piston returns too much, the driver will notice a lengthening of the pedal travel.

Bleed screws are used for bleeding the system and to eliminate air bubbles in the caliper.

15.5.2 Friction Material

To achieve the maximum performances of a braking system for race applications, the friction material has to guarantee both high friction and high stability. During long deceleration, disk and pad temperatures can rise dramatically, and if the friction material changes its characteristics with temperature, the modulation of the braking force can become very difficult. In Figure 15.11,

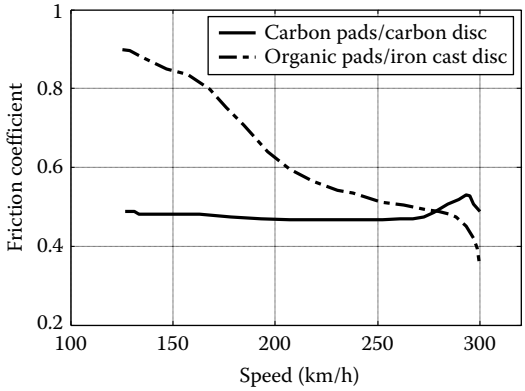


FIGURE 15.11 Comparison of friction properties as a function of speed between an organic pad coupled with a cast iron disk (dot–dash line) and a carbon pad coupled with a carbon disk (solid line).

the tendencies of the friction coefficient as a function of speed during a braking maneuver from 300 to 120 km/h are reported for the two main kinds of friction materials, namely, organic or semisintered pads coupled with cast iron disk and carbon pads coupled with carbon disk. From the figure, it can be noted the very high stability of the carbon pads, which also show a higher values of friction coefficient at the very beginning of the maneuver. This early stage of the braking is very important because in this phase the braking system is usually not able to lock the wheels and the driver can exert all its force to achieve the maximum possible deceleration. A sharp speed decrement at the beginning of the maneuver is important for reducing the braking space.

Another advantage of carbon pads and carbon disk is the lower density of carbon with respect to cast iron (about five times less), which allow for the production of much lighter brakes, thus reducing the unsprung mass of the vehicle. Moreover, being carbon a composite material, the fiber arrangement can be designed in order to give preferential conductivity direction.

In comparison with organic or semisintered pads coupled with cast iron, carbon pads coupled with carbon disk are up to ten times more expensive.

15.6 CONCLUSIONS

In this chapter, the main issues concerning the braking system of a racing car are described. At first, the simplified equations for the computation of the braking forces and the braking balance of a car with aerodynamic surfaces are derived.

Then, the brake balance of a racing car is discussed. In this case, the ideal brake balance is the function of the actual speed and varies during braking. An adaptive balance regulator should be employed for optimizing the braking performances; however, most racing regulations allow only constant brake balancing, so a compromise has to be found between higher and lower speeds. The braking system of an F1 car is described empathizing the devices for the regulation of the brake balance.

Finally, starting from the computation of the dissipated energy during braking and the required peak braking power, the main requirements in the design of the braking system are defined, with particular attention to braking calipers and pads.

The information reported in this chapter allows the interested reader to understand the basics of race car dynamics during braking. A designer can use this approach as a starting point for braking performance calculations.

REFERENCES

1. G. Genta, *Meccanica Dell'autoveicolo*, Levrotto & Bella, Turin, Italy, 2000.
2. T. Gillipsie, *Fundamentals of Vehicle Dynamics*, SAE, Warrendale, PA, 1992.
3. C. Doniselli, G. Mastinu, and M. Gobbi, Aerodynamic effects on ride comfort and road holding of automobiles, *Proceedings of the XIV IAVSD Symposium*, Swets & Zeitlinger, Ann Arbor, MI, 1995.
4. E. Benzing, *Dall'aerodinamica alla Potenza in Formula 1*, Giorgio Nada Editore, Vimodrone, Italy, 2004, ISBN 88-7911-318-6.
5. B. Breuer and K. H. Bill, *Brake Technology Handbook*, SAE, Warrendale, PA, 2008.
6. W. F. Milliken and D. L. Milliken, *Race Car Vehicle Dynamics*, SAE International, Warrendale, PA, 1995, ISBN 1-56091-526-9.

16 Aerodynamics and Vehicle Dynamics

Andreas Wagner

CONTENTS

16.1 Introduction 529
16.2 Definition of Aerodynamic Forces and Moments 530
16.3 Occurrence of Aerodynamic Forces and Moments 531
16.3.1 Airflow around the Vehicle 531
16.3.2 Drag and Lift with Different Rear End Shapes 534
16.3.3 Induced Drag (Vortex Drag) 536
16.4 Determination of Aerodynamic Forces and Moments 538
16.4.1 Lift Behavior 540
16.5 Crosswind Behavior 543
16.5.1 Vehicle Excitation by Crosswind 543
16.5.2 Vehicle Reactions to Crosswind 545
16.5.3 Evaluation of Crosswind Behavior 547
16.6 Dynamics of Vehicle Excitation 552
16.7 Conclusion 554
Symbols 555
References 556

16.1 INTRODUCTION

Particular attention has already been paid to the interfaces between aerodynamics and vehicle dynamics. This is increasingly important in the modern development process. Knowledge of both technical areas is required when a vehicle is in the development stage. The aim of this chapter is to provide a basic understanding of these interfaces.

First, an explanation will be given on how aerodynamic forces arise and how they interact with each other. It will be shown how aerodynamic forces depend on the design of the basic shape of the vehicle, particularly the shape of the rear end. The relationship between lift and drag assumes a central role. That illustrates what types of conflicts arise between goals in aerodynamic development.

Furthermore, it will be shown how aerodynamic forces and moments affect the behavior of the vehicle. One central topic in this regard is the effect of lift, while another is the vehicle's response to crosswind. A study of the crosswind behavior in particular requires a comprehensive consideration of the driver–vehicle system with its aerodynamic and chassis characteristics. Finally, aerodynamic effects on the vehicle in unsteady flow conditions will be discussed.

This chapter deliberately avoids going into the measures that can be adopted to optimize the shape of the vehicle with respect to, for example, reducing drag. The aim is rather to address aerodynamic influences and its interfaces with vehicle dynamics.

16.2 DEFINITION OF AERODYNAMIC FORCES AND MOMENTS

The air flowing around the vehicle causes a resulting force and a resulting moment. Figure 16.1 shows them broken down into three forces and three moments in a coordinate system specific to the vehicle. A position centrally located along the wheelbase is selected as the reference point. The lift force F_L and the pitch moment M_y as well as the lateral force F_S and the yaw moment M_z can each be substituted by forces acting at the ground level on the wheels. By using this definition, the aerodynamic forces at the front and rear axles can be considered. The substitute forces are given by the following relationships:

Lift:

$$F_L = F_{Lf} + F_{Lr} \quad M_y = (F_{Lf} - F_{Lr}) \cdot \frac{l_0}{2}$$

Lateral force:

$$F_S = F_{Sf} + F_{Sr} \quad M_Z = (F_{Sf} - F_{Sr}) \cdot \frac{l_0}{2}$$

Forces and moments are not stated in absolute terms in aerodynamics. The so-called aerodynamic coefficients are used instead. The most readily recognized is the drag coefficient c_D, often shortened to "c_D value." The derivation of the drag coefficient is given by Equation 16.1. The numerator can be interpreted as the longitudinal force, which would arise, if the entire kinetic energy of the flow was imparted to the vehicle. The denominator equals the measured drag. The ratio of both produces a dimensionless parameter that describes the quality of the shape from an aerodynamic point of view. The c_D value is (largely) independent of the flow velocity:

$$c_D = \frac{F_D}{(\rho/2) \cdot A \cdot v^2} \tag{16.1}$$

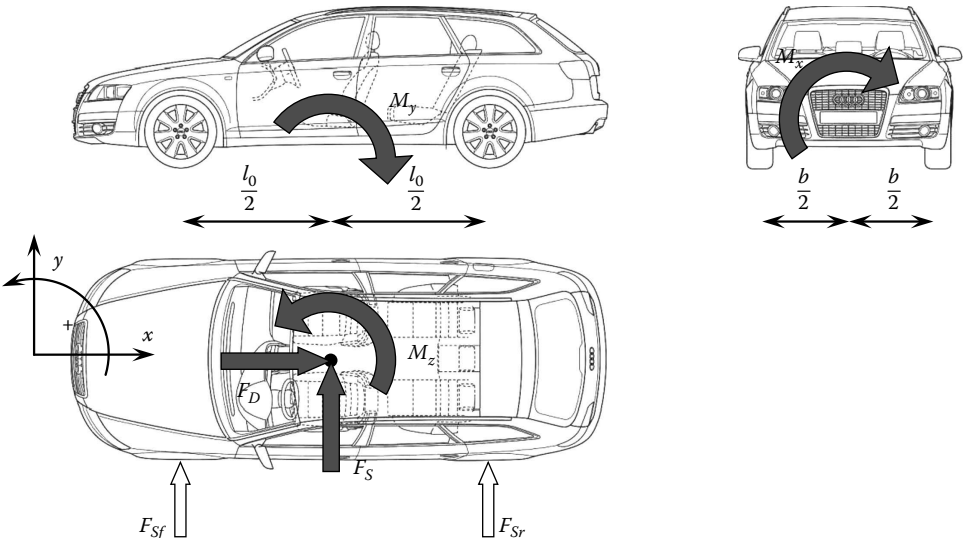


FIGURE 16.1 Definition of aerodynamic forces and moments.

The other aerodynamic coefficients are formed in a similar fashion to Equation 16.1. The aerodynamic coefficients of a vehicle are measured in wind tunnels (see Section 16.4). The following equations are used to calculate the aerodynamic forces and moments:

Drag force:

$$F_D = \frac{\rho}{2} \cdot c_D \cdot A \cdot v^2 \tag{16.2}$$

Lateral force:

$$F_S = \frac{\rho}{2} \cdot c_S \cdot A \cdot v^2 = \frac{\rho}{2} \cdot (c_{Sf} + c_{Sr}) \cdot A \cdot v^2 \tag{16.3}$$

Lift:

$$F_L = \frac{\rho}{2} \cdot c_L \cdot A \cdot v^2 = \frac{\rho}{2} \cdot (c_{Lf} + c_{Lr}) \cdot A \cdot v^2 \tag{16.4}$$

Roll moment:

$$M_x = \frac{\rho}{2} \cdot c_{Mx} \cdot l_0 \cdot A \cdot v^2 \tag{16.5}$$

Pitch moment:

$$M_y = \frac{\rho}{2} \cdot c_{My} \cdot l_0 \cdot A \cdot v^2 \tag{16.6}$$

Yaw moment:

$$M_z = \frac{\rho}{2} \cdot c_{Mz} \cdot l_0 \cdot A \cdot v^2 \tag{16.7}$$

The aerodynamic coefficients depend on the vehicle's angle of incidence (angle of attack). Figure 16.2 shows the aerodynamic coefficients and their dependency on the angle of attack for a production vehicle referring to a notchback as an example.

In addition to engine power and gear ratios, drag is a contributory factor determining top speed and fuel consumption of a vehicle. Lift forces and the pitch moment affect vehicle dynamics. The lateral force and the roll and yaw moments are of particular interest when looking at unsteady flow conditions, such as crosswind.

16.3 OCCURRENCE OF AERODYNAMIC FORCES AND MOMENTS

16.3.1 Airflow around the Vehicle

As a vehicle moves, the surrounding air must diverge around the body of the vehicle. As a result, accelerations and decelerations occur in the surrounding airflow. These deviations can be made visible by introducing smoke into the airstream in front of the vehicle (Figure 16.3). The so-called

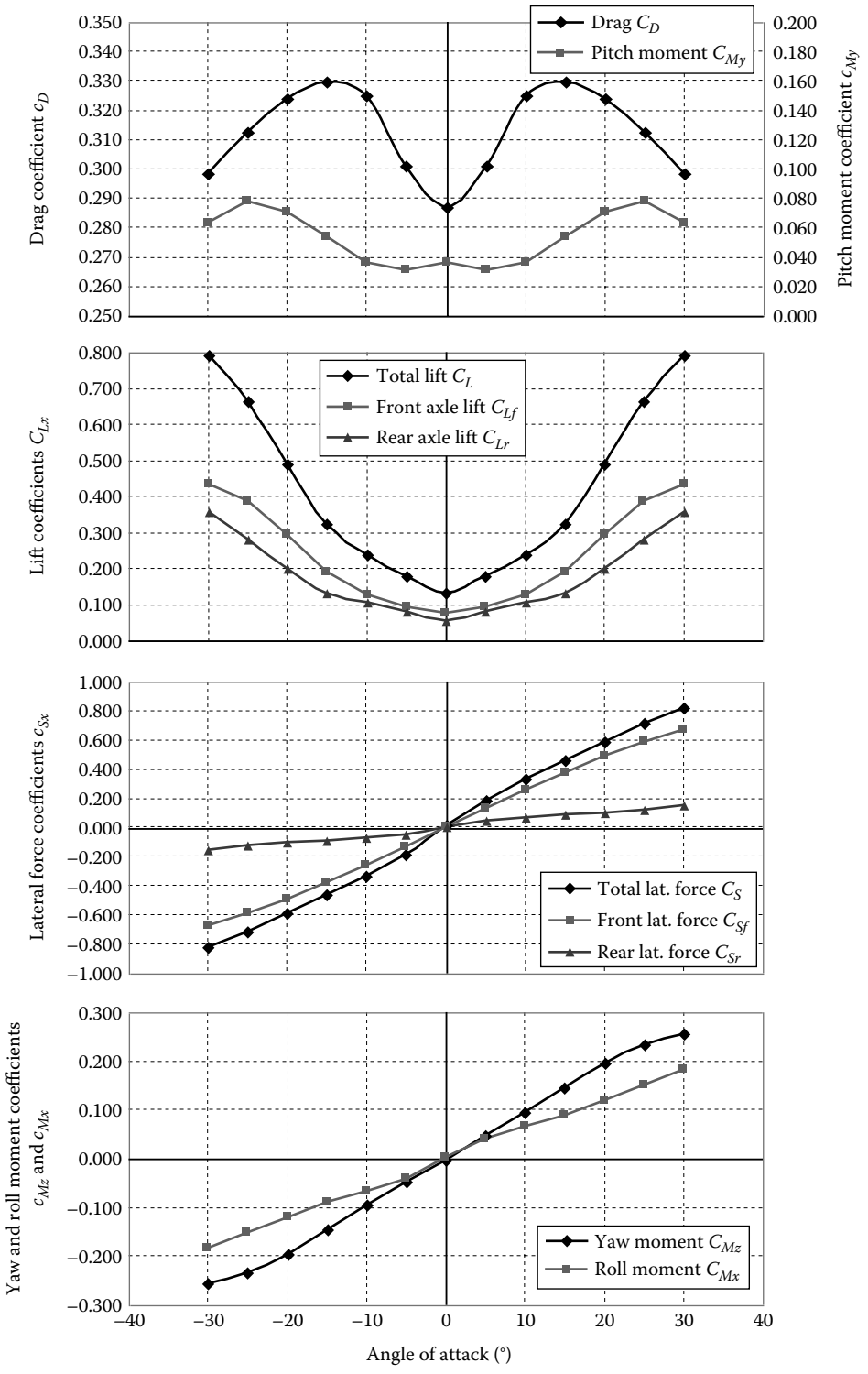


FIGURE 16.2 Aerodynamic coefficients for a production car (notchback). (Courtesy of Audi AG, Ingolstadt, Germany.)



FIGURE 16.3 Airflow around a vehicle. (Courtesy of Audi AG, Ingolstadt, Germany.)

streamlines become visible, which are bent corresponding to the vehicle's shape. In areas where the streamlines are compressed, the flow accelerates. Regions with widened streamlines indicate decelerations in the airflow. If frictionless flow is considered, the physical relationships within the airflow can be described by the Bernoulli equation [1]:

$$\frac{p}{\rho} + \frac{v^2}{2} = const. \tag{16.8}$$

Simple but nevertheless important relationships can be derived using this equation. It states clearly that the energy contained in the airflow is stored in the form of pressure work and kinetic energy (velocity). Between both forms of energy, an exchange can take place. High flow velocities lead to low pressures in the airflow and vice versa. Considering the airflow around the vehicle in Figure 16.3, the pressure decreases when the flow accelerates (compression of the streamlines), whereas positive pressure gradients are produced where the flow decelerates (widening of the streamlines). This energy exchange results in a pressure distribution over the outer skin of the vehicle, which is shown diagrammatically in Figure 16.4 for a notchback vehicle.

Forces and moments can be determined by integrating the pressures over the vehicle's skin. In the case of attached flow, the pressure distribution is produced by accelerations and decelerations in

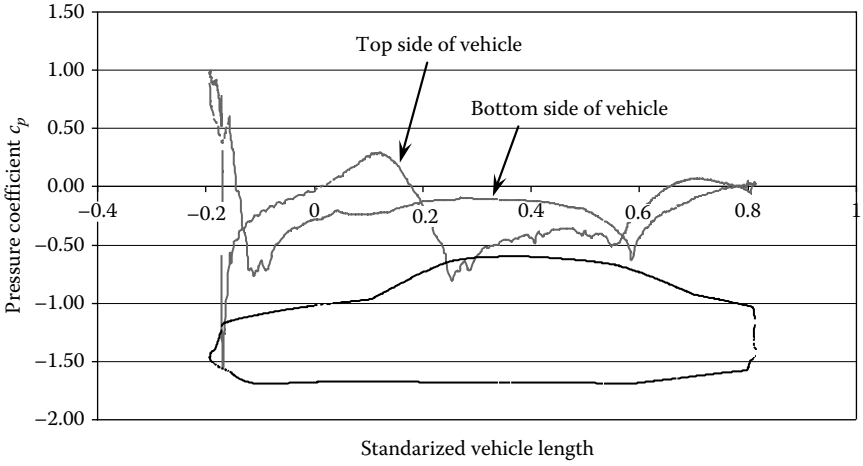


FIGURE 16.4 Pressures induced by an airflow around a vehicle. (Courtesy of FKFS, Stuttgart, Germany.)

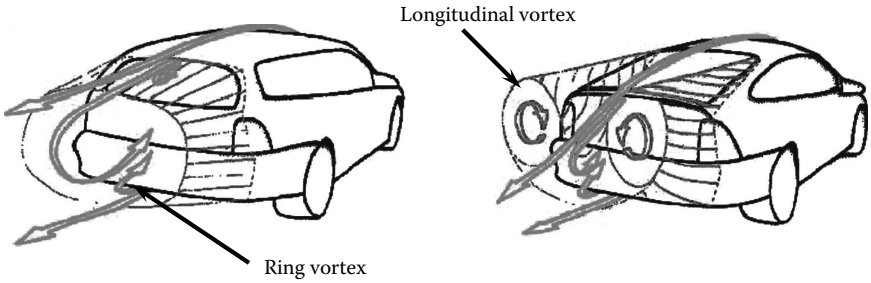


FIGURE 16.5 Types of vortex in a vehicle's wake. (From Howell, J., Shape and drag, in *Using Aerodynamics to Improve the Properties of Cars*, International Short Course on Vehicle Aerodynamics at FKFS, Stuttgart, Germany, 1998.)

the airflow. In the area of the vehicle's rear end, the flow deviates sharply, which causes high positive pressure gradients (widening of streamlines in Figure 16.3). As a result of these positive pressure gradients, the flow separates and forms a separation bubble behind the vehicle. This so-called wake is marked by pronounced vortices in the flow. Vortices are high energy flow forms containing not only rotational energy but also frictional losses. They generally result in low-pressure regions. The Bernoulli equation (16.8) is no longer suitable to describe their physical state. This type of flow is described by the so-called Navier–Stokes equations [1–4].

Due to high kinetic energies and friction losses, the pressures inside a vortex are low. In addition, vortices induce high flow velocities in their surrounding, which finally lead to low-pressure regions. The origination and formation of vortices at the rear of a vehicle therefore affect, to a large extent, the dynamic forces and moments. Two types of vortex can be observed in a vehicle's wake (Figure 16.5).

The first type is created at sharp edges at the rear of the vehicle that runs perpendicular to the flow direction (such as the rear edge of the trunk lid). These vortices rotate around an axis that is parallel to the separation line and form the vehicle's wake. For distinction purposes, this type of vortex will be called "ring vortex." The second type of vortex formation is created at edges running obliquely to the direction of flow. The flow separates along these edges and begins to form tube-shaped horseshoe vortices with rotational axes in the direction of flow. They continue to roll further on this axis along the flow direction. Their existence is not limited to the vehicle's wake. These vortices occur in pairs at the rear of the vehicle and cause a so-called downwash in the vehicle's wake. High flow rates predominate in this downwash, resulting in low pressures.

In particular, the longitudinal vortices create regions with very low pressures and, as a result, have a great effect on drag and lift. Both types of vortex (longitudinal and ring vortices) can be found with any vehicle. However, depending on the shape of the vehicle, one or the other of these types is prevalent. The intensity of the particular vortex type is determined by the shape of the rear. This will be explained in the next section.

16.3.2 Drag and Lift with Different Rear End Shapes

In aerodynamic terms, a distinction has to be made between three rear end shapes—squareback, notchback, and fastback. Each of these rear end shapes is associated with a characteristic form of vortices (cf. [1,5–7]). The following illustrations will clarify this aspect.

Figure 16.6 shows the flow past the rear of a squareback. The separation edges at the rear run more or less perpendicular to the flow direction. Ring vortices are formed that constitute the vehicle's wake. As a result, low pressures predominate in the wake. Due to their geometry, squareback vehicles, compared with other rear shapes, have a large wake. That is, in conjunction with the comparatively large contact surfaces, the low pressures cause a higher drag. The type of flow at the rear end of the vehicle also affects the lift forces in particular at the rear axle. Typical for this type of rear flow pattern is a tendency to negative lift or at least very low levels of lift at the rear axle.

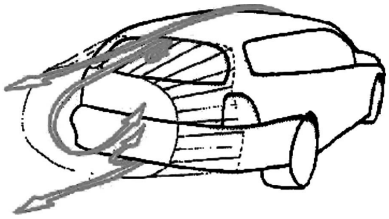


FIGURE 16.6 Diagram illustrating the flow leaving the rear of a squareback vehicle. (From Howell, J., Shape and drag, in *Using Aerodynamics to Improve the Properties of Cars*, International Short Course on Vehicle Aerodynamics at FKFS, Stuttgart, Germany, 1998.)

Figure 16.7 shows a diagram of the flow past the rear end of a notchback. It has mainly the same type of separation at the rear of the vehicle as with a squareback. Ring vortices are formed that run perpendicular to the separation edges. However, the flow partially starts to separate at the *C* pillars (see Chapter 26). Longitudinal vortices roll away, rotating inward and causing a downwash in the wake. Notchbacks therefore have a tendency to rear lift in comparison to squarebacks. Overall, the effect of the ring vortices predominates in the wake. Since the wake of a notchback is comparatively smaller than in a squareback, it has less drag.

Figure 16.8 illustrates the flow down the rear end of a fastback. The flow separates at the sloping edges of the *C* pillars and begins to form longitudinal vortices that rotate inward, dominating the separation pattern. They create a strong downwash behind the vehicle. Furthermore, they continue to roll downstream and have still not dissipated some distance behind the vehicle. The longitudinal vortices create low-pressure regions that, compared with the other two rear end shapes, are linked to high drag coupled with high lift values.

Whether ring vortices or longitudinal vortices predominate in a vehicle's wake is largely affected by the slant angle of the rear end shape. The longitudinal vortices form at slanting edges, such as at the *C* pillars. This explains why this type of vortex predominates particularly in fastback shapes. This phenomenon was investigated and studied in detail in [1], where the slant angle of a vehicle's

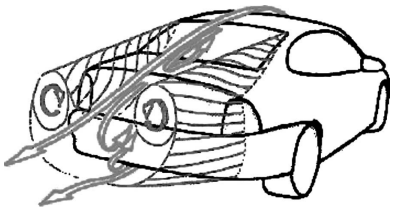


FIGURE 16.7 Diagram illustrating the flow leaving the rear of a notchback vehicle. (From Howell, J., Shape and drag, in *Using Aerodynamics to Improve the Properties of Cars*, International Short Course on Vehicle Aerodynamics at FKFS, Stuttgart, Germany, 1998.)

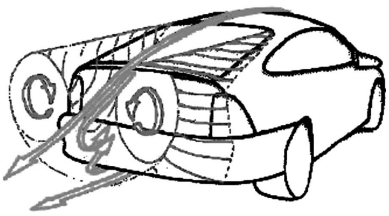


FIGURE 16.8 Diagram illustrating the flow leaving the rear of a fastback vehicle. (From Howell, J., Shape and drag, in *Using Aerodynamics to Improve the Properties of Cars*, International Short Course on Vehicle Aerodynamics at FKFS, Stuttgart, Germany, 1998.)

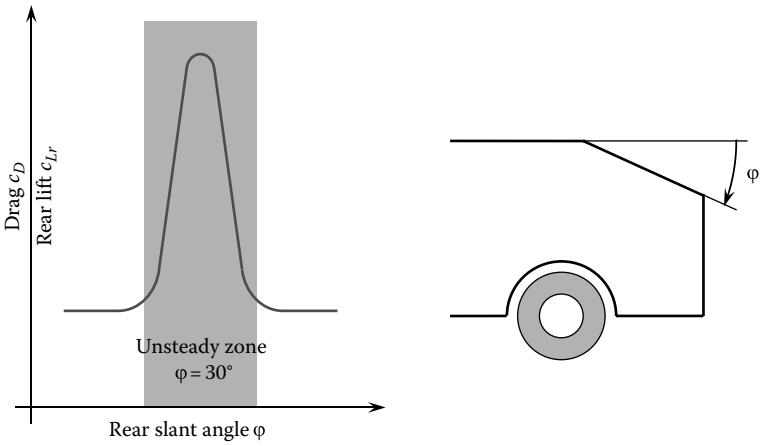


FIGURE 16.9 Effect of slant angle on drag and lift.

rear end was varied in order to observe the effect on lift and drag. Figure 16.9 illustrates the results schematically. The formation of longitudinal vortices reaches a maximum with a slant angle of about 30°. The wake of the vehicle has a marked downwash effect, induced by the inwardly rotating longitudinal vortices. Drag and lift increase significantly in this range.

The quoted slant angle of about 30° was determined on an idealized basic vehicle body. Detailed elements in the design of a rear end, such as radii and geometric tapering of the surfaces, partially diminish the described effects so that this critical angle is subject to scatter within certain limits [1,5].

The quoted example is intended to show that both drag and lift are already determined by the basic shape of the vehicle within a certain window. Rear end shapes with sloping separation edges, particularly fastbacks, favor the formation of longitudinal vortices consuming high energy. Even though significantly less pronounced, those longitudinal vortices separating from the rear of a notchback should be noted. Both rear end shapes exhibit rear lift, depending on the extent of the longitudinal vortices. In squareback vehicles, the separation at the rear is dominated by the formation of ring vortices. This type of rear flow pattern has a tendency to have negative rear lift or at least very low levels of rear lift.

Whether the rear end shape produces positive or negative rear lift is also reflected in the direction of rotation of the vortices behind the vehicle. Notchbacks and fastbacks tend to have positive rear lift. In other words, a higher pressure acts on the underside of the vehicle than on the top of the vehicle. Thus, the pressures equalize at the rear end from "bottom to top." Looking from behind, the vortices rotate inward. The effect is opposite in the case of squarebacks.

16.3.3 Induced Drag (Vortex Drag)

The examples in the previous section show that drag and lift are always influenced by the occurrence of vortices. This is termed the so-called induced drag (cf. [8]). The term induced drag stems originally from aerospace and will be explained below by means of an airfoil.

Figure 16.10 illustrates how drag occurs for an airfoil when lift is created. The drag on the airfoil consists of two components. First, the airfoil itself has its own drag. This is caused by friction effects in the flow that manifest themselves as pressure losses. Secondly, lift is created by the airfoil caused by positive pressure on the underside of the wing (pressure side) and low pressures on the upper side (suction side). This can be produced by either a setting angle of the airfoil or an appropriate shaping of the airfoil profile. Due to the difference in pressure between the pressure and suction sides, pressure has to be equalized at the wing tips. That creates edge vortices that roll off behind the wing, developing into a so-called horseshoe vortex and causing a downwash behind the airfoil. The component of drag that must be used to create this pair of edge vortices is known as induced drag.

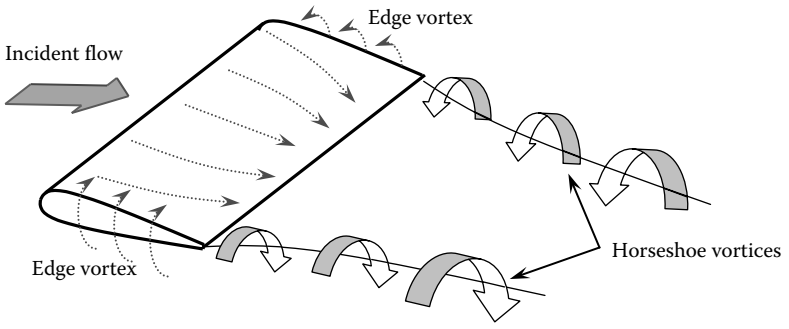


FIGURE 16.10 Illustration of the flow from an airfoil.

The induced drag depends quadratically on the generated lift force, that is, $c_{D,ind} \sim c_L^2$. Considering the airfoil in Figure 16.10, drag is at its lowest when neither positive nor negative lift is created, thus producing no induced drag.

This relationship can be read across, as a general rule, to earth-bound vehicles (cf. [9–12]). Instead of the term "induced drag," "vortex drag" is used in vehicle aerodynamics. Due to the interaction between the vehicle and the road surface, rotating wheels and the cooling air intake flow, etc., the minimum drag does not occur at zero lift but rather at low lift values ($0 < c_{ah} < 0.1$). The creation of positive or negative lift on a vehicle influences vortex formation and drag. In other words, drag and lift depend fundamentally to each other.

By using this model of induced drag or vortex drag, some aerodynamic effects can be better understood. In conjunction with Section 16.3.2, it can be again illustrated why different rear end shapes, due to their design, have lower drag or lower positive or negative lift. Figure 16.11 illustrates the quadratic relationship between drag and lift. As explained in Section 16.3.2, in the case of fastbacks, a pair of longitudinal vortices is formed that causes high lift. The result is high vortex drag and therefore high drag values. Longitudinal vortices are generally not dominant in notchbacks and squarebacks, and they therefore experience little vortex drag. The creation of negative lift in racing cars by adding components such as airfoils or spoilers can therefore also result in high vortex drag. For the sake of completeness, it should be mentioned at this point, however, that racing cars in addition use the so-called ground effects. That is, they achieve their negative lift with only little induced drag by utilizing the interaction between the underside of the vehicle and the road surface (cf. [1]).

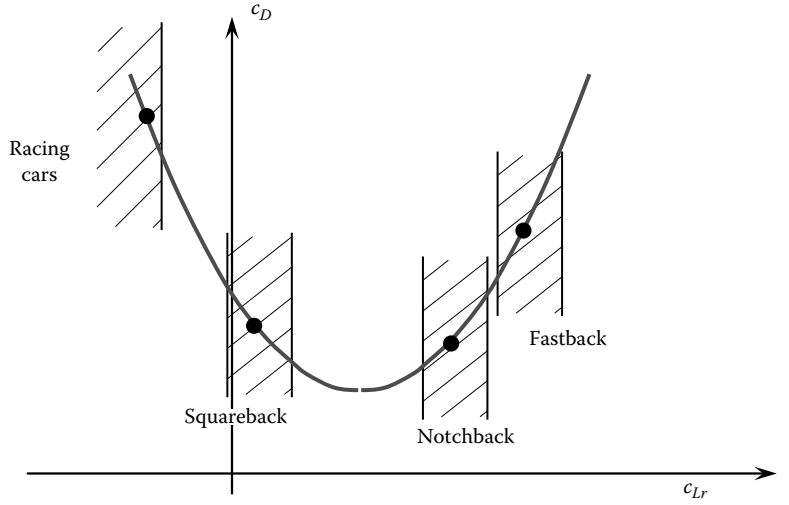


FIGURE 16.11 "Classification" of rear end shapes according to vortex drag.

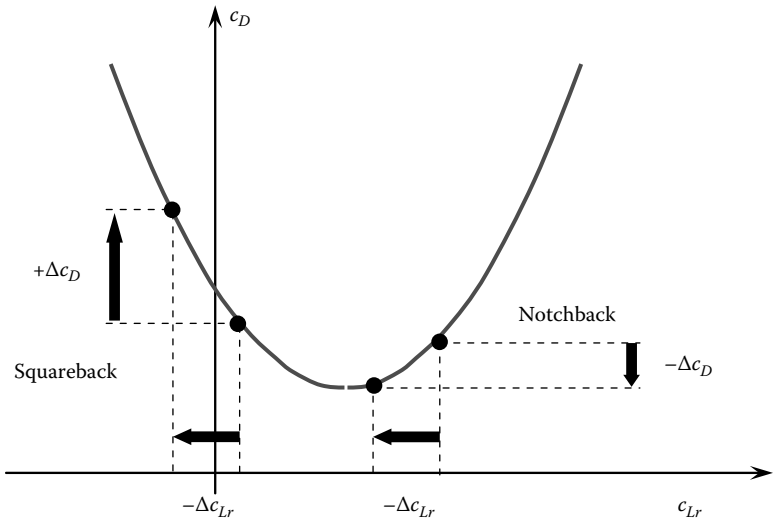


FIGURE 16.12 Relationship between drag and rear axle lift. (From Wagner, A. et al., Einfluss der Bodensimulation auf den induzierten Widerstand von Fahrzeugen. *Proceedings of the 6th International Symposium on Automotive and Engine Technology*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Stuttgart, Germany, 2005.)

Utilizing this interpretation of vortex drag helps also to understand why individual rear end shapes must be treated differently to some extent in aerodynamic development [11,12]. As an example, in Figure 16.12, the intention is to fit a rear spoiler to a notchback and a squareback vehicle to generate negative lift. This spoiler is to be incorporated fully into the bodywork to avoid the possibility of having separate vortex systems from the spoiler and from the basic vehicle shape. Due to their lift relationships, notchback vehicles are usually located to the right of the vortex drag parabola's minimum. If the lift of this type of vehicle is reduced by, for example, incorporating a spoiler, the vortex drag reduces and therefore the vehicle's drag also. Since squareback vehicles tend to exhibit negative lift, they are located on the left side of the parabola's minimum. If a spoiler is incorporated in a vehicle with a squareback to produce negative lift, the vortex drag will increase.

This example is meant to illustrate that aerodynamic measures used on a variety of rear end shapes can result in opposite effects on drag.

The aim of aerodynamic development is to create low drag as well as low lift forces. This is achieved, on the one hand, by optimizing a vehicle's design and styling (cf. [1,7,13–17]). On the other hand, vortices induce low-pressure regions and, thus, increase drag and lift. The creation of vortices, particularly the formation of longitudinal vortices at the rear of the vehicle, should be avoided. It must be noted in this context that different rear end shapes, due to the characteristic way in which the airflow separates from their rear end, have built-in aerodynamic characteristics regarding drag and lift that is inherent in their design and styling (see Section 16.3.2). Due to vortex drag, drag and lift cannot be influenced separately from each other.

16.4 DETERMINATION OF AERODYNAMIC FORCES AND MOMENTS

The aerodynamic development of vehicles and the determination of their aerodynamic forces and moments are carried out in wind tunnels (cf. [18–21]). Figure 16.13 shows a typical design of a wind tunnel. The air is accelerated through a fan and through the nozzle into the measurement section, called "plenum." After flowing around the vehicle, the air is captured by the collector and recirculated to the fan. The vehicle is located in the plenum on a balance that determines the aerodynamic forces and moments.

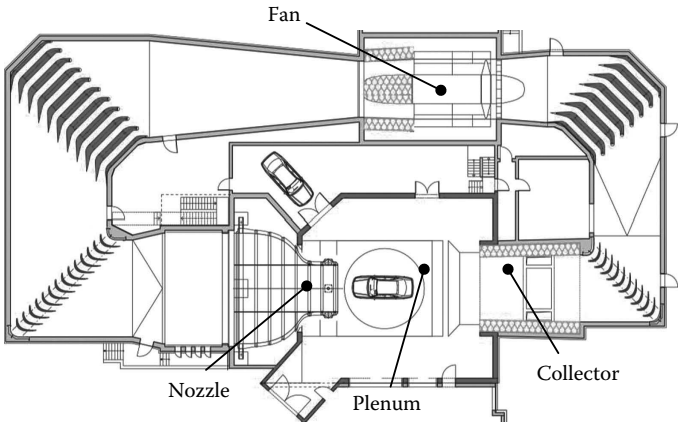


FIGURE 16.13 Layout of a wind tunnel for vehicles. (Courtesy of Audi AG, Ingolstadt, Germany.)

The flow in wind tunnels is an approximation to the flow conditions on road. The quality of this approximation depends on the technical design of the wind tunnel (cf. [22,23]). In real life, a vehicle moves in an "infinitely" large flow field, while the measurement section in a wind tunnel is spatially limited. As a result, the measurement section is blocked by the presence of a vehicle, which means that the air jet has to spread out to flow around the vehicle and to be received by the collector. As a result, pressure gradients are formed as well as accelerations and decelerations that influence the results of the aerodynamic forces depending on the way the wind tunnel was designed. Engineers endeavor to design a wind tunnel in order to minimize these types of effect, to compensate for them, and/or to balance one against the other (cf. [24–26]).

Another approximation to conditions on road has to be seen in the fact that in a wind tunnel, the air is blown around a stationary vehicle, while in real life the vehicle travels through the air. In newly designed wind tunnels, it is therefore possible to apply a so-called ground simulation consisting of several components [27–29] (Figure 16.14). Suction and blowing devices are used to affect the boundary layer existing due to the relative motion between the air and the floor of the wind tunnel. A moving belt operates between the vehicle's tracks to simulate the relative motion between the vehicle and the road surface. The wheels are rotated at the appropriate speed by small moving belts. Thus, the vehicle no longer stands on the balance with the tire's contact patch but instead must be

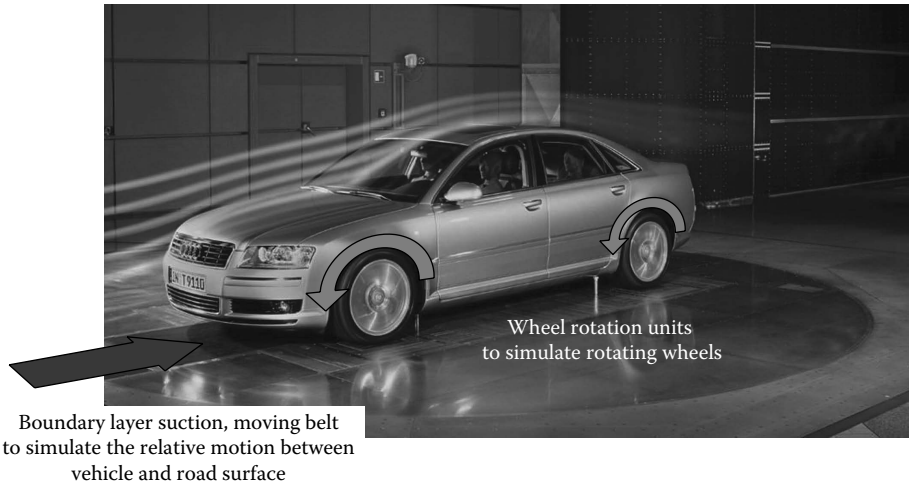


FIGURE 16.14 Ground simulation in the Audi wind tunnel. (Courtesy of Audi AG, Ingolstadt, Germany.)

held using special retainers, on the rockers, for example. The use of ground simulation provides a more precise reproduction of the relationships in actual travel on road [30–34].

Numerical flow simulation (computational fluid dynamics [CFD]) provides a second method of determining aerodynamic forces and moments [1,3,4,35]. The Navier–Stokes equations are used in CFD for a numerical resolution of the flow field. In this method, a number of modeling concepts have to be used, such as modeling the turbulent shear stresses that occur due to the high-frequency, small-scale, turbulent fluctuations in the flow. This so-called turbulence modeling is based on simplified assumptions and represents an approximation to actual physical conditions (cf. [4,36,37]). Both can result in fundamental discrepancies between the actual and simulated flows.

As computing power has increased, the development of CFD has accelerated in the last few years. Using modern methods, it is possible to obtain a particularly good correlation with actual flow conditions. As a result, the use of CFD has become an indispensable tool in vehicle development today: both in the early development stage of a vehicle, when a physical model is not yet available, as well as in later development stages to accompany wind tunnel measurements. The most important benefit of CFD, however, is the possibility to a detailed insight into the flow (cf. [2,38–41]).

16.4.1 Lift Behavior

From a dynamic motion point of view, lift forces on a vehicle are very important, since they can affect both intrinsic steering behavior and braking behavior and, therefore, the ability to control the vehicle (cf. [42–46]). A well-balanced design of vehicle dynamics is reflected in a slightly understeering behavior (see Chapter 27). This requires a targeted and coordinated design of axles, steering system, elasto-kinematics, etc., as well as weight distribution. As shown in Section 16.2, the lift forces increase quadratically with the speed of the incident flow and superimpose on the wheel loads resulting from the weight distribution. This reduces the downforces acting through the wheels dependent on the road speed.

To discuss the results on vehicle dynamics, among others, the tire has to be considered. Tire characteristics are explained in detail in Chapters 17 through 20 and 27. In this chapter, some basic diagrams are illustrated as follows.

Figure 16.15 shows the force distribution in a vehicle traveling in a right turn. For the sake of illustration, only the left-hand side of the vehicle is to be considered. The centrifugal force F_Z resulting from the curved path requires a lateral reactive force F_S, which acts at road surface level and which has to be applied by the tires. Furthermore, due to the centrifugal force F_Z, a shift in wheel load takes place from inside to outside of the curve. This component $\Delta F_{N,dyn}$ superimposes on the

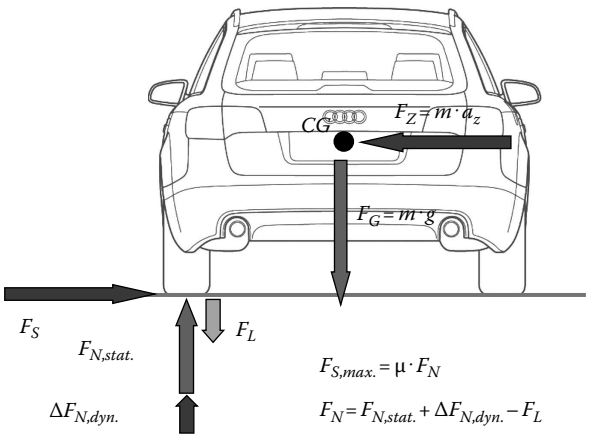


FIGURE 16.15 Force distribution in a vehicle performing a right turn with lift effect (forces at the left side of the vehicle are shown only).

static component of the wheel load $F_{N,stat}$, which results from the weight of the vehicle F_G. The wheel load is increased on the outer side of the curve due to this component. The ability of the tires to apply a lateral reactive force depends on their normal force. As the normal force F_N increases, the tires can apply higher maximum lateral forces $F_{S,max}$.

If a vehicle is subject to lift, the load on the tires is reduced by the lift force F_L. The centrifugal force F_Z and the vehicle weight F_G as well as the resulting reactive forces (F_S, $F_{N,stat}$, $\Delta F_{N,dyn}$) remain unchanged since they depend on the mass of the vehicle. To be able to perform the turn at the same lateral acceleration in Figure 16.15, the tires must apply the same lateral force when lift is present. The normal force F_N is reduced, however, by the lift force F_L. As a result, the ability of the tires to apply a lateral reaction is reduced. In other words, the maximum lateral forces $F_{S,max}$ they can apply are lowered. The result is a decrease in the achievable lateral acceleration a_Z and thus lower speeds when turning.

Figure 16.15 shows the effect of lift, looking at the vehicle's outer side of the curve. To show the effect of lift on the intrinsic steering behavior of a vehicle, it is necessary to consider the lift forces on the front and rear axles separately. The following characteristic diagram for a tire will be a starting point.

Figure 16.16 diagrams the relationship between normal force, sideslip angle, and transferable lateral force for a tire. As the lateral force increases, the sideslip angle increases. With small lateral forces, the tire behavior is approximately linear, that is, a doubling of the lateral force results in a doubling of the sideslip angle. As the lateral force increases further, the tire behaves regressively. Once the maximum transferable lateral force is reached, the tire begins to skid. The sideslip angle increases and higher lateral forces can no longer be sustained. The influence of the normal force on the tire's lateral control force was already mentioned before: an increase of the normal force leads to greater cornering forces and vice versa.

If the lift forces reduce the tire's normal force, the same lateral force can be applied only at greater sideslip angles (see Figure 16.16). As a result, the lateral stiffness of the axle in question decreases. Lift on the rear axle produces a tendency to oversteer, while lift on the front axle results in a tendency to understeer. The opposite is the case for negative lift.

Similar effects are produced in the case of fluctuating lift forces. Figure 16.17 illustrates the dependency of the transferable lateral force of a tire on its normal force. The relationship is regressive since the transferable lateral force cannot be increased to the same extent as the normal force. The normal force is now allowed to fluctuate periodically by varying the lift force around a mean value $F_{N,0}$. This oscillation of the normal force results in a fluctuation in the lateral control force around a mean value $F_{S,0}$. Due to the regressive behavior, the mean value of the lateral control force $F_{S,0}$ is less than the lateral force $F_{S,stat}$, which was assigned to the mean value of the wheel load. Thus, when the normal force fluctuates, the tire can apply less lateral force on the average. However, to be able to counter the lateral forces present, a greater sideslip angle needs to be applied to the tire. Oscillating lift forces on the rear axle alter the intrinsic steering behavior toward oversteering. Oscillating lift forces on the front axle result in a tendency to understeer.

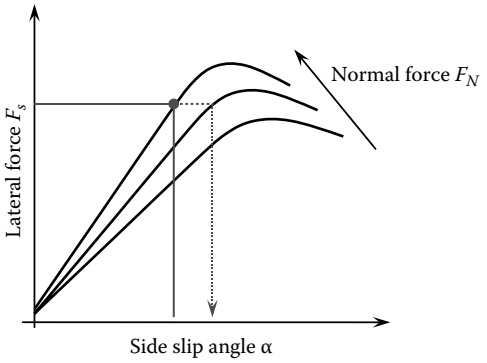


FIGURE 16.16 Relationship between lateral force, normal force, and sideslip angle of a tire.

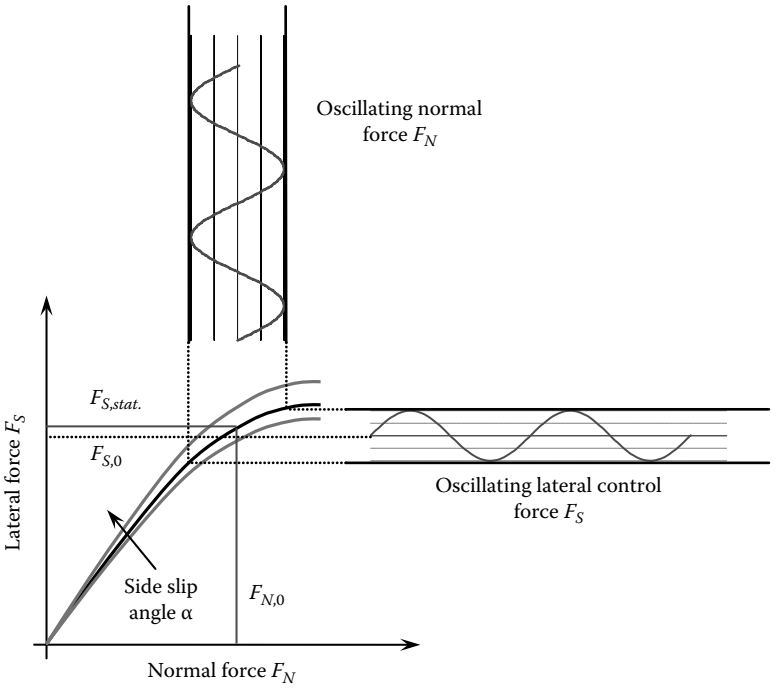


FIGURE 16.17 Effect of an oscillating normal force on the lateral control ability of a tire.

These two examples illustrate the effect of lift forces on the intrinsic steering behavior. Lift forces result in lower normal forces on the tire. When lift is present, the affected axle behaves more elastic in lateral direction. As a result, an increased sideslip angle is required to counter the lateral forces. The opposite happens in the case of negative lift. Negative lift on the rear axle causes an understeering behavior. Negative lift on the front axle results in an oversteering tendency but has a positive effect on steering behavior. High downward pressures on the front axle produce a direct steering sensation and improve the response behavior of the vehicle to steering inputs (e.g., peak response time [Equation 16.16]).

Another aspect of lift is its effect on braking (see also Chapter 15). Figure 16.18 illustrates the force distribution of a braking vehicle affected by lift. The position of the center of gravity (*CG)* and the vehicle mass m causes static normal forces $F_{Nf,stat}$ on the front axle and $F_{Nr,stat}$ on the rear axle. The deceleration from the brake application results in a dynamic wheel load shift from the rear to

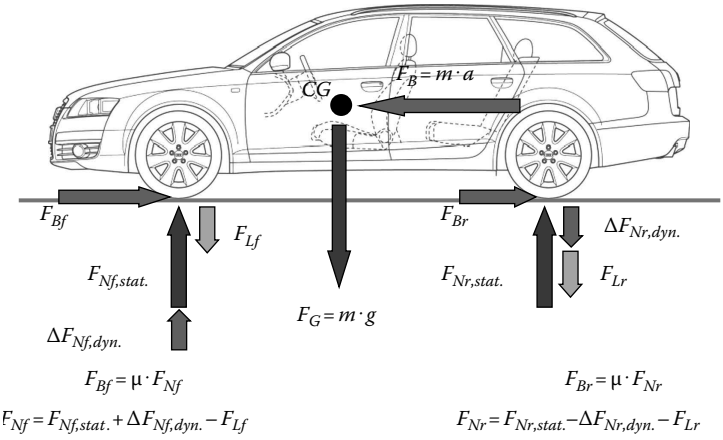


FIGURE 16.18 Force distribution in a vehicle when braking with lift present.

the front axle. The normal force on the front axle is increased by the dynamic component $\Delta F_{Nf,dyn}$ and reduced on the rear axle by the component $\Delta F_{Nr,dyn}$.

In order to be able to achieve maximum braking deceleration, the road grip at the front and rear axles must be exploited uniformly. Thus, the distribution of the braking forces to the front and rear axles F_{Bf} and F_{Br} must be in the same relationship as the normal forces on the front and rear axles F_{Nf} and F_{Nr}. This distribution is known as "ideal brake force distribution" (see also [42]) and is permanently installed in the vehicle for a defined brake deceleration (0.8 g) by designing the brake system accordingly.

If a vehicle is subject to lift, the normal forces on the axle reduce due to the superimposed lift forces. As a result, the maximum braking forces that the tire can apply are reduced, as is the maximum attainable braking deceleration. Furthermore, the ratio of the normal forces at the front and rear axles can be affected by lift in a way that it no longer corresponds to the design criteria for ideal braking force distribution. Rear lift reduces the maximum transferable braking force at the rear axle without affecting the front axle. As a consequence, the road grip at the rear axle is utilized more strongly, which results in a tendency to lock the rear axle with hard braking. Since lift forces, as explained in Section 16.2, increase quadratically with speed, this phenomenon is more pronounced when braking at high speeds.

This tendency, which the rear axle locks first, must be avoided for safety reasons by designing the braking system appropriately. The distribution of the braking forces between the front and rear axles must therefore be designed for braking at high speed, when the rear is subject to maximum lift forces. However, in the case of braking at lower speeds, that is, lower lift forces at the rear, the vehicle is not utilizing the road grip ideally and cannot attain the maximum achievable deceleration.

In racing cars, the aerodynamic development of high negative lift forces to achieve high lateral control forces and therefore high attainable speeds in curves is of great importance (see Chapter 15). For production vehicles, it is desirable to produce only little positive or negative lift to maintain the vehicle dynamics properties over the widest possible range of road speeds. Since lift forces increase quadratically with speed, an unequally decreasing aerodynamic load to the front or the rear axle can lead to a tendency to over- or understeer as a function of driving speed. At the same time, this will also affect braking. The aerodynamic balance, that is, the ratio of front and rear lift, should correspond to the weight distribution. Thus, the load on both axles would be applied or relieved uniformly, that is, lift forces would not change the ratio of the normal forces. In certain circumstances, this can represent a conflict in objectives with the aerodynamic optimization of drag (cf. Section 16.3.3).

16.5 CROSSWIND BEHAVIOR

16.5.1 Vehicle Excitation by Crosswind

Figure 16.19 shows the incident airflow relationships when crosswind is present. The "vehicle motion" airstream v_{veh} flows around the vehicle at approximately the travel speed. A crosswind component with a wind speed of v_{Wind} approaching at an angle of τ_{Wind} is superimposed on the

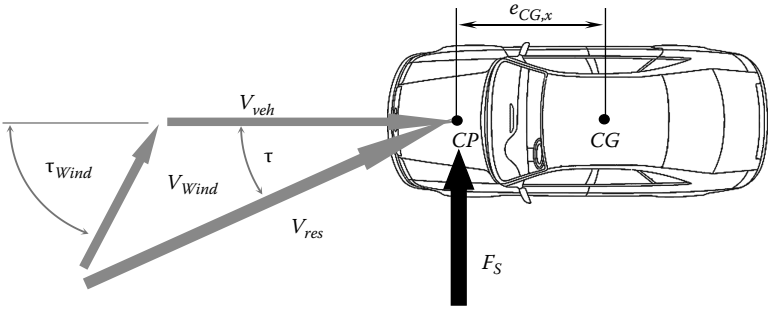


FIGURE 16.19 Resulting flow meeting the vehicle. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

motion airstream. The vector addition of both components produces the resulting airflow meeting the vehicle at a speed of v_{res} at a resulting angle of attack τ:

$$F_S = \frac{\rho}{2} \cdot v_{res}^2 \cdot A \cdot c_S \approx \frac{\rho}{2} \cdot A \cdot k_S \cdot \tau \cdot v_{res}^2 = K \cdot \tau \cdot v_{res}^2 \quad (\text{for } \tau < 15°) \tag{16.9}$$

Equation 16.9 is used to consider the side force created due to crosswind. The aerodynamic lateral force coefficient can be assumed to be linear, to a good approximation, within a range for angles of attack of ±15°. This applies, as a general rule, to production vehicles with different rear end shapes. By assuming a linear relationship for the aerodynamic lateral force coefficient $c_S = k_S\tau$, the crosswind excitation can be expressed by the product τv_{res}^2. This product depends on the wind values v_{Wind} and τ_{Wind} and can be used to express the intensity of crosswind. Normally, the product τv_{res}^2 is used as a measure of the wind disturbance and as a reference value when considering the vehicle's response behavior, such as by means of a magnitude ratio [47,48].

Crosswind induces a yaw moment M_Z as well as a lateral wind-force F_S. The yaw moment can be taken into account, to a reasonable approximation, as the lateral wind-force F_S acts through the nominal center of pressure *CP* with a lever arm $e_{CG,x}$. This lever arm can be determined from the aerodynamic coefficients (Equations 16.3 and 16.7) and is described by means of Equation 16.10 in dependency of the angle of attack. The same procedure applies to consideration of the roll moment, again to a reasonable approximation (not shown in Figure 16.19):

$$e_{CG,x} = \frac{c_{Mz} \cdot l_0}{c_S} + \frac{l_0}{2} - l_r \tag{16.10}$$

A linear, planar, single-track model, acted on by the lateral wind-force in Equation 16.9 (Figure 16.20), will be considered to discuss the vehicle's crosswind behavior. The derivation of the equations of

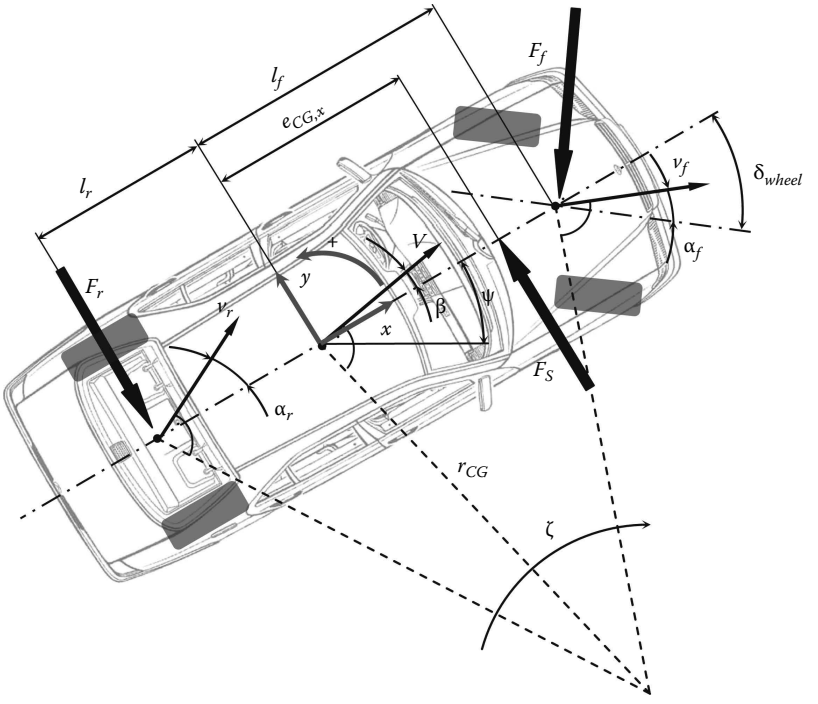


FIGURE 16.20 Single-track model.

motion will not be introduced at this point, and the reader is referred to the literature [42,43,45,46,49]. The equation system 16.11 describes the response of the single-track model to steering inputs and to crosswind excitation. Equations 16.12 through 16.16 show the most important system-related values for this single-track model:

$$\begin{bmatrix} \ddot{\psi} \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} -\dfrac{c_r \cdot l_r^2 - c_f \cdot l_f^2}{J_z \cdot v} & \dfrac{c_r \cdot l_r - c_f \cdot l_f}{J_z} \\ \dfrac{c_r \cdot l_r - c_f \cdot l_f - m \cdot v^2}{m \cdot v^2} & -\dfrac{c_r + c_f}{m \cdot v} \end{bmatrix} \cdot \begin{bmatrix} \dot{\psi} \\ \beta \end{bmatrix} + \begin{bmatrix} \dfrac{(\rho/2) \cdot A \cdot k_S \cdot e_{CG,x}}{J_z} \\ \dfrac{(\rho/2) \cdot A \cdot k_S}{m \cdot v} \end{bmatrix} \cdot \tau v_{res}^2 + \begin{bmatrix} \dfrac{c_f \cdot l_f}{J_z \cdot i_L} \\ \dfrac{c_f}{m \cdot v \cdot i_L} \end{bmatrix} \cdot \delta_L \tag{16.11}$$

Undamped natural angular frequency:

$$\omega_0 = \sqrt{\frac{m \cdot v^2 \cdot (c_r \cdot l_r - c_f \cdot l_f) + c_r \cdot c_f \cdot (l_r + l_f)^2}{m \cdot v^2 \cdot J_z}} \tag{16.12}$$

Damping time constant:

$$\delta = \frac{m \cdot (c_r \cdot l_r^2 + c_f \cdot l_f^2) + J \cdot (c_f + c_r)}{2 \cdot m \cdot v \cdot J_z} \tag{16.13}$$

Damped natural angular frequency:

$$\omega_d = \sqrt{\omega_0^2 - \delta^2} \tag{16.14}$$

Damping factor:

$$\vartheta = \frac{\delta}{\omega_0} \tag{16.15}$$

Peak response time of yaw rate:

$$T_{\dot{\psi},\max} = \frac{1}{\omega_d} \cdot \arctan \frac{\omega_d}{\delta - (c_r \cdot (l_f + l_r)/m \cdot v \cdot l_f)} \tag{16.16}$$

16.5.2 Vehicle Reactions to Crosswind

Lateral forces may occur in normal travel when a vehicle is met by a diagonal airflow, as is common in windy conditions. They cause the vehicle to deviate from the desired course (such as driving in a straight line) and take the form of lateral and yaw responses. Typical situations consciously registered by the driver are, for example, driving across bridges or overtaking a truck at high speed, as well as—in a sufficiently strong stochastic wind—traveling on open stretches. The driver has to react with steering adjustments to correct for the wind gusts. Depending on the sensitivity of the vehicle to crosswind, this is primarily discomforting but can also become a safety issue.

Due to their stochastic nature, the frequency domain is used to describe crosswind effects on a vehicle. To "normalize" the response of the vehicle to the crosswind excitation, it can be represented by the magnitude ratio [50]. Figure 16.21 shows an example of a magnitude ratio of the yaw velocity with wind excitation at different travel speeds. A constant crosswind produces a constant yaw velocity (magnitude ratio with a frequency $f = 0$ Hz). Due to the decreased damping of the vehicle at high driving speeds, the resonance magnitudes of the yaw velocity increase as the speed increases.

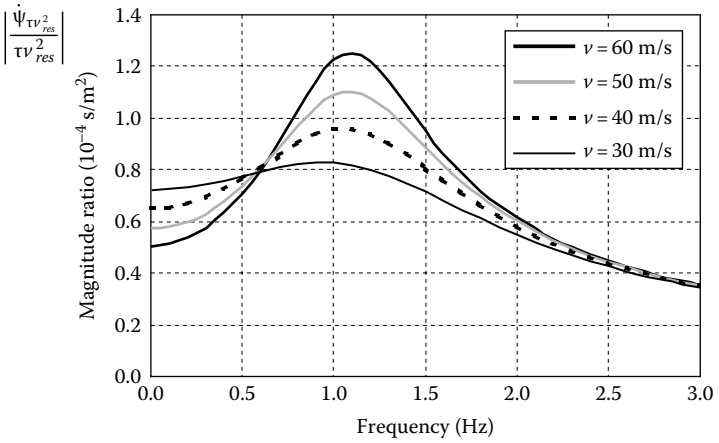


FIGURE 16.21 Magnitude ratio and phase shift of the yaw velocity with crosswind excitation at different travel speeds. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

All other vehicle reactions display a similar response to crosswind excitation as shown in Figure 16.21 for the yaw velocity (see [51]). The crosswind disturbance appears to the driver as a pattern consisting of these reactions. The composition of this pattern depends on the vehicle design and its characteristics and is therefore typical for the vehicle being driven [52]. This vehicle reaction pattern causes a swerve, requiring a steering adjustment by the driver.

Yaw velocity is an important motion quantity in this reaction pattern since it results in a perceptible angular deviation from the course. Crosswind behavior is often considered using the yaw velocity reaction only. When looking at the resonance behavior of the vehicle in a crosswind, however, the vehicle reaction pattern must be considered as a whole, in other words, taking into account all other vehicle reactions.

Figure 16.22 shows the dependency of the damped natural frequency and resonance frequencies on the travel speed with crosswind excitation. The axes were arranged in a way that the abscissa

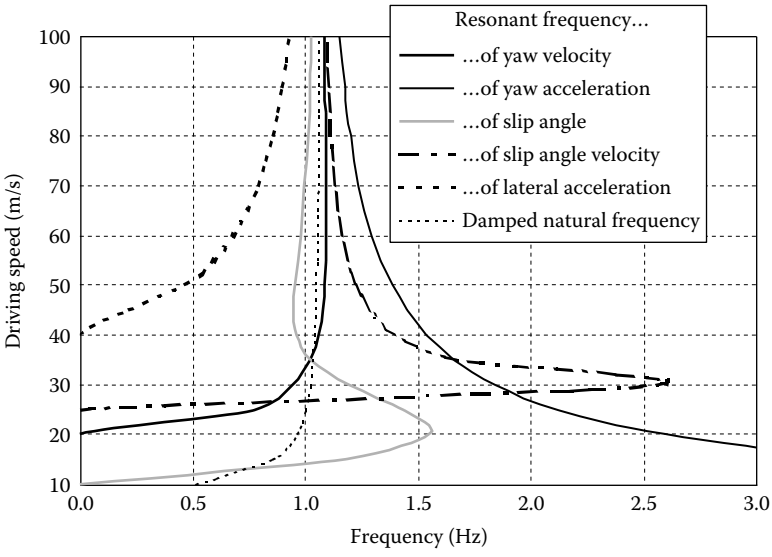


FIGURE 16.22 Dependency of the resonant frequencies of a vehicle on the vehicle's speed with crosswind excitation. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

would represent the frequency as in Figure 16.21. The resonant frequencies are scattered over a wide range within a realistic speed range up to about 50 m/s. At further increases in speed, they approach the damped natural frequency of the vehicle.

If the wind gusts occur with time intervals between each other, the vehicle is allowed to oscillate after excitation. These decaying oscillations occur at a natural frequency determined by the system, in this case the damped natural frequency of the vehicle. When the vehicle is allowed to naturally oscillate, all vehicle reactions vibrate at this frequency. The vehicle's yaw velocity reaction is very important in this case, since the damped natural frequency of the vehicle runs close to the resonant frequency of the yaw velocity. This kind of excitation is typical for passing across bridges or passing trucks. The driver generally understands the interference he or she can expect in these situations and often makes an initial steering adjustment before the interference occurs. These kinds of situations are classed as less critical [51].

If the vehicle is excited by stochastic crosswind, a permanent stochastic vehicle excitation superimposes on the decaying natural oscillations. In contrast to the "sporadic" vehicle excitation by individual wind gusts as described earlier, in this case, the vehicle is excited over the whole range of its resonant frequencies. As a result, the individual vehicle reactions resonate alternately. The yaw velocity reaction, again, is important due to the proximity of its resonant frequency to the undamped natural frequency. Only considering the yaw velocity, however, in this case is not sufficient. All vehicle reactions must be taken into account when stochastic excitation of the vehicle occurs.

16.5.3 Evaluation of Crosswind Behavior

The weaker the excitation by crosswind is, the lower the aerodynamic lateral force and the yaw moment are. It is an obvious objective, therefore, to keep both as low as possible by targeted aerodynamic development. Since crosswind effects mainly involve comfort problems, the subjective perception by the driver is an important factor. It is therefore far more complicated to obtain a clear definition whether a vehicle is sensitive to crosswind using objective criteria (see [53]). This will be explained in more detail in the succeeding text.

To assess the susceptibility of a vehicle to crosswind disturbances, the first step can be to consider the tendency to yaw (Figure 16.23). Due to their rear end shape, squareback vehicles

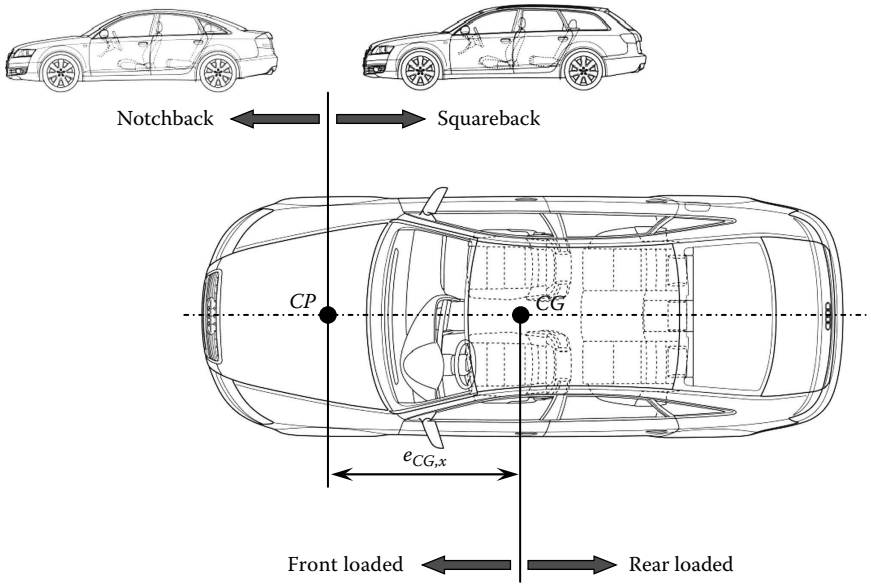


FIGURE 16.23 Estimation of the wind susceptibility using the yaw tendency.

in comparison to notchbacks have more contact surface area at their rear against which a crosswind gust can act. Their *CP* is further back and is therefore closer to the *CG* than it is the case with notchbacks. As a rule, squarebacks therefore experience greater lateral forces but lower yaw moments. In other words, squareback vehicles tend to have less strong yaw reactions to crosswind. The same applies to front-loaded vehicles. If the vehicle's *CG* is further forward, the reactions to crosswind are less intense due to the shortened lever arm $e_{CG,x}$ of the lateral wind-force.

This estimation of yaw reactions enables a first statement to be made with regard to the "crosswind susceptibility" of a vehicle from a stationary point of view. If statements are to be made about a vehicle's time-dependent behavior in crosswinds, other approaches need to be pursued.

One of these approaches is passing a crosswind facility with the vehicle [54,55], which has been used widely up to now. In such a facility, the vehicle is driven through an air jet artificially produced by one or more blowers (Figure 16.24). The measurements can be run in an "open-loop" mode, that is, with locked steering, or in "closed-loop" mode, that is, with compensatory steering inputs by the driver. The assessment takes place in the time domain.

In open-loop maneuvers, the gradient, the maximum value, and the overshoot of the vehicle reactions when passing through the wind facility as well as the diminishing oscillations of the vehicle as it exits are evaluated. The aim is to minimize the vehicle reactions and their overshoots, particularly with regard to yaw reactions. A residual yaw angle and a track deviation remain as the vehicle exits the facility. Similar to Figure 16.23, a statement can be drawn from these quantities whether a vehicle tends more strongly toward yaw or lateral reactions. In closed-loop maneuvers, the driver actions to counter the wind effect as well as the vehicle reactions to both wind gust and steering intervention are evaluated. The evaluation criteria and target functions are specific to each manufacturer.

The evaluation of crosswind behavior by passing through crosswind facilities is disputed [47,51,56]: The flow profile of the artificially produced crosswind differs from natural conditions

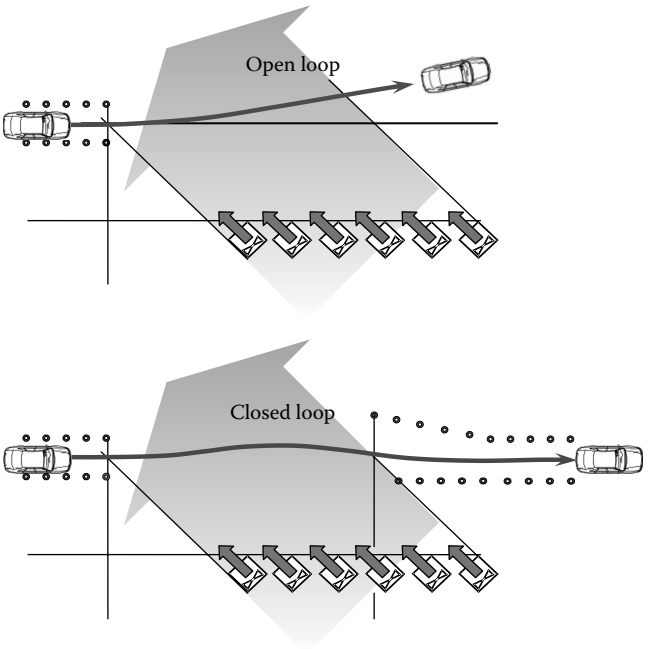


FIGURE 16.24 Illustration of crosswind-facility maneuvers in (a) open loop—fixed/free control—and (b) closed loop. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

that can have a different effect on a vehicle. In closed-loop maneuvers, the driver actions cannot be evaluated realistically since he or she is prepared to react to the wind gust. The driver's subjective perception in natural and artificial wind is often shown to be barely adequate.

The situation simulated in a crosswind facility is analogous to crossing over a bridge or overtaking a truck. The study of a vehicle's behavior in stochastic crosswind has been gaining favor, particularly over the last few years. The vehicle's response has been studied increasingly in the frequency domain to evaluate the crosswind behavior. This should give a better correlation between the objective evaluation and the driver's subjective impression of the vehicle's behavior.

In [47,48], the crosswind behavior was evaluated in natural, stochastic wind by driving on public roads. The magnitude ratio of the yaw velocity to the wind excitation was evaluated (see Figure 16.21). The aim was to minimize the resonant magnitudes of the yaw velocity. In [48], the vehicle reactions to both crosswind and driver's actions, that is, the overall reaction, were observed. In [47], the vehicle reactions exclusive to crosswind, that is, without the driver's actions, were assessed. To eliminate the effect of the driver from the measurements on public roads, the vehicle reactions to steering interventions were determined and "subtracted" from the measured vehicle reactions using single-track models.

Even this method does not produce a complete correlation between objective evaluations and subjective driving impressions [51]. This is due to the fact that the vehicle reactions exclusive to crosswind are considered and minimized. Some vehicles are perceived by the driver to be sensitive to crosswind, although they objectively display low reactions to stochastic wind and vice versa. To reach a complete correlation with the subjective assessment, the driver has to be included in the evaluation.

In [51], a procedure was developed to include the driver in the evaluation as a central member of the driver–vehicle system. The driver was observed in the driver–vehicle control loop during on-road measurements as illustrated in Figure 16.25. The wind disturbance causes vehicle reactions (index τv_{res}^2), which are perceived by the driver and corrected for with steering interventions. The vehicle reactions due to the driver's actions (index δ_L) superimpose on the vehicle reactions due to crosswind (index τv_{res}^2) to form overall vehicle reactions. These overall reactions serve as information to the driver on how effective his or her steering interventions were to compensate for the crosswind gust. Equation 16.18 describes this relationship mathematically.

In Figure 16.25, the driver acts as a "sensor" in terms of his or her perception of vehicle reactions but also as an "actuator" in view of his or her compensatory steering actions. It may occur that steering corrections made by the driver can impair a vehicle's behavior to crosswind (cf. [57]).

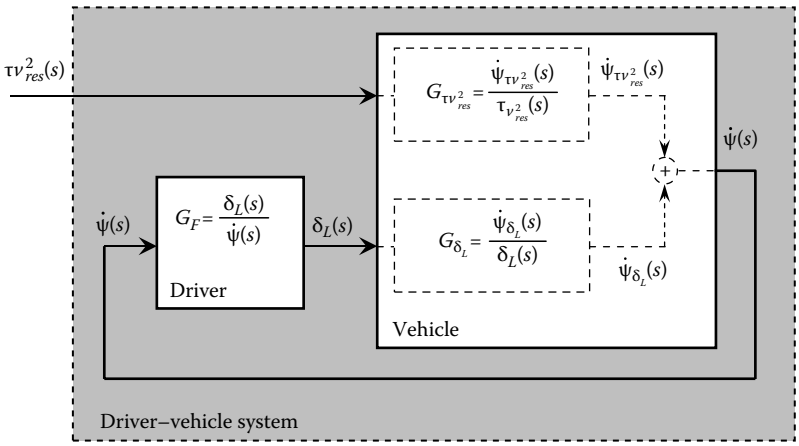


FIGURE 16.25 Signal flow diagram for the driver–vehicle system. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

On the other hand, a vehicle's behavior objectively assessed to be negative can be improved by the steering actions of the driver. For this reason, instead of concentrating exclusively on the direct response of the vehicle to the crosswind excitation, optimizing the interaction between driver and vehicle promises to be more effective.

This is achieved by evaluating the characteristic driver's steering intervention in terms of how he or she processes the information from the vehicle reactions [44]. All vehicle reactions to crosswind appear to the driver as a reaction pattern. During crosswind disturbances, the driver is able, even with different vehicles, to maintain a narrow and similar path. To do this, he or she has to adapt his or her steering interventions to the vehicle's reaction pattern. That is, he or she changes his or her reaction to the information. This change is reflected in the characteristic of his or her compensatory steering input and evaluated by the mentioned approach as illustrated as follows:

$$\frac{\dot{\psi}}{\tau v_{res}^2}(j\omega) = \frac{\dot{\psi}_{\tau v_{res}^2}}{\tau v_{res}^2}(j\omega) + \left[\frac{\dot{\psi}_{\delta_L}}{\delta_L}(j\omega) \cdot \frac{\delta_L}{\tau v_{res}^2}(j\omega) \right] \tag{16.17}$$

Figure 16.26 shows a magnitude ratio of the driver's compensatory steering input related to the wind disturbance and their phase shift from a measurement under stochastic crosswind (driver's actions; Equation 16.17). Using the magnitude ratio, it is possible to analyze how strongly the driver reacts to the individual vehicle reactions [51]. A strong driver reaction to the so-called vestibular information, such as yaw velocity and yaw acceleration, is expressed in high magnitude ratios at frequencies

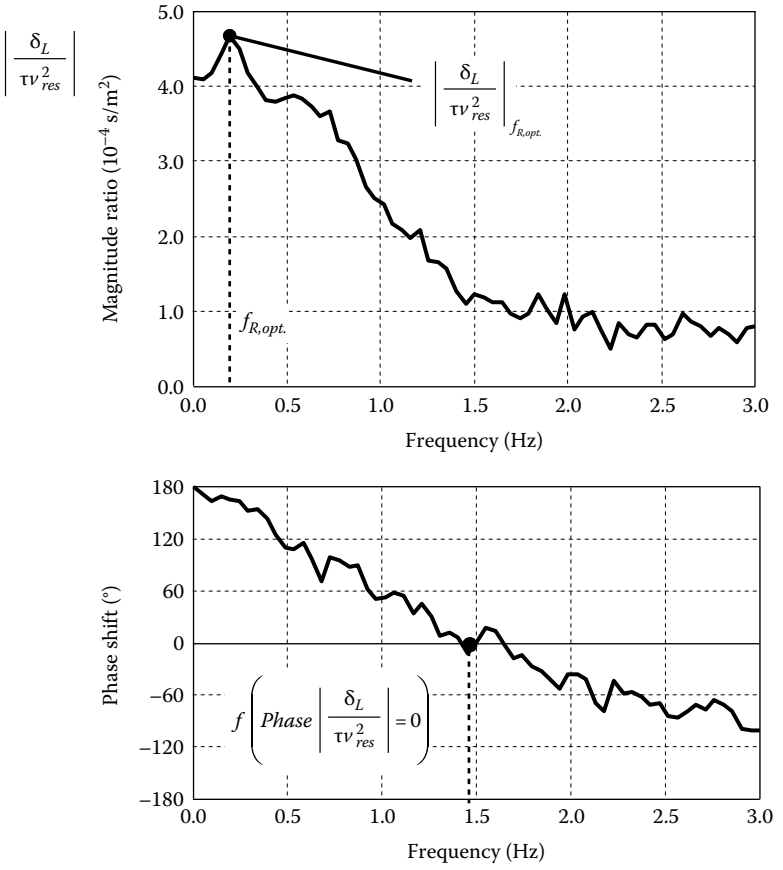


FIGURE 16.26 Magnitude ratio and phase shift of the steering angle to the wind disturbance. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

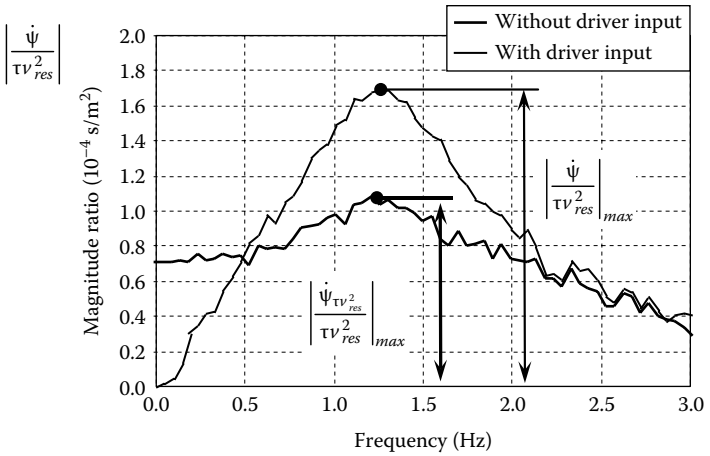


FIGURE 16.27 Magnitude ratio of the yaw velocity with a crosswind disturbance, with and without driver input. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

between 0.8 and 1.5 Hz and in flat phase-shift gradients. The phase shift therefore becomes an indicator of how spontaneously and how fast the driver reacts. Steep gradients of the phase shift result from slow driver reactions and vice versa. Strong driver reactions to visual information such as track deviation or yaw angle are reflected in high magnitude ratios in a frequency range from 0 to about 0.6 Hz.

The effectiveness of this driver action is indicated by the magnitude ratio of the vehicle reactions to the wind disturbance with and without driver input (Figure 16.27). The vehicle reactions without driver input are produced exclusively by the crosswind excitation. The vehicle reaction with driver input is produced by both wind excitation and driver's actions to compensate for the wind excitation (overall vehicle reactions). In a frequency range of about 0.5–2.0 Hz, the driver intensifies the vehicle reactions to the wind by his or her steering interventions. This is typical for all vehicles and results, on the one hand, from the driver's reaction time and, on the other, from the phase shifts of the vehicle reactions [51].

In [51], four criteria expressing the driver's subjective assessment are defined from Figures 16.26 and 16.27:

- If a driver reacts strongly to visual information, a high magnitude ratio is produced in Figure 16.26 in the visual information processing region (0–0.6 Hz). Moreover, the maximum is shifted to higher frequencies. Vehicles in which the driver reacts with this characteristic are felt subjectively to be "indirect" and "requiring steering effort." The evaluation of the amount of steering effort is given by Equation 16.18. High values denote a negative assessment:

$$K_L = \left|\frac{\delta_L}{\tau v_{res}^2}\right|_{f_{R,opt}} \cdot f_{R,opt} \tag{16.18}$$

- Strong reactions to vestibular information result in flat gradients of the phase shift between steering angle and wind excitation in Figure 16.26. These flat gradients reflect fast driver reactions. Vehicles that cause this characteristic driver reaction are felt to be "twitchy" and "unpredictable." This aspect can be evaluated using the frequency at which the phase shift passes through zero (Equation 16.19). High values denote a negative assessment:

$$K_P = f\left(\text{Phase}\left(\frac{\delta_L}{\tau v_{res}^2}\right) = 0\right) \tag{16.19}$$

- The extent to which the driver amplifies the vehicle's reactions by his or her steering interventions (Figure 16.27) can be expressed by the ratio of the resonant magnitudes of the yaw velocity with and without driver input. Equation 16.20 defines a so-called amplification factor. In the case of high amplification factors, the driver perceives the wind disturbance as "difficult to counter":

$$K_V = \frac{\left|\dot{\psi}/\tau v_{res}^2\right|_{\max}}{\left|\dot{\psi}_{\tau v_{res}^2}/\tau v_{res}^2\right|_{\max}} \tag{16.20}$$

- The resonance magnitude of the yaw velocity is defined in Equation 16.21 with driver input (Figure 16.27) as a further criterion. In the case of high resonance magnitudes with driver's actions (overall vehicle reactions), the driver perceives the vehicle reactions as "severe":

$$K_G = \left|\frac{\dot{\psi}}{\tau v_{res}^2}\right|_{\max} \tag{16.21}$$

It must be noted that the perceived intensity of the crosswind excitation is reflected in the vehicle reactions with driver's actions (overall vehicle reactions), which are due to both crosswind and driver's steering actions. It could be shown in [51] that the driver does not directly detect the vehicle reactions exclusively due to crosswind. Earlier procedures, however, aimed at minimizing the vehicle reactions exclusively due to crosswind. This explains why it was not sufficient in earlier procedures to observe the vehicle only.

The individual criteria represented by Equations 16.18 through 16.21 can be combined in one overall evaluation (Equation 16.22), in which different weightings are applied to each of the criteria. High numbers of the overall evaluation denote negative overall assessments. This procedure takes into consideration how the subjective impression during driving occurs: a driver evaluates the crosswind behavior of a vehicle from a number of individual factors and not only based on the reactions of the vehicle:

$$K_{ges} = 2\cdot[K_L] + 5\cdot[K_P] + 3\cdot[K_V] + 1\cdot[K_G] \tag{16.22}$$

Whether the evaluation of the crosswind behavior of a vehicle is positive depends to a large extent on how the driver reacts to the wind disturbance in the vehicle being tested. This procedure optimizes the crosswind behavior of a vehicle by means of improving the interaction between driver and vehicle. In arriving at this definition, both vehicle dynamics and aerodynamics are considered since both affect the vehicle reaction pattern and therefore the characteristics of the driver's actions.

The previously described approach was developed using a wide range of vehicles and the results checked with respect to the influence of several drivers and road speeds. The results corresponded in detail with the subjective perceptions of the drivers. However, performing measurements under natural crosswind is often related with huge time exposures. This was why, in [51], a virtual driver was modeled who is able to adapt the characteristics of his or her steering interventions to different vehicles realistically (for driver models, see Chapter 38). It is possible, utilizing this virtual driver, to use the previously described evaluation procedure also in simulations on the whole virtual driver–vehicle system. Thus, the interaction between driver and vehicle can be studied virtually in an early stage of vehicle development.

16.6 DYNAMICS OF VEHICLE EXCITATION

It is necessary, at this point, to mention one aspect of aerodynamics that occurs in cases of unsteady vehicle excitation. The way a vehicle is excited by crosswind is subject to dynamic processes (cf. [58–62]). This is explained in the following text using a sudden occurrence of a lateral wind gust as an example.

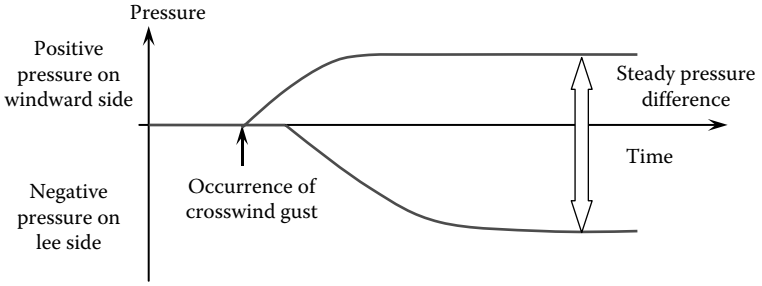


FIGURE 16.28 Principle of pressure buildups on windward and lee sides.

Crosswind forces arise when a vehicle is met by an air jet from the side, in which positive pressure occurs overall on the side facing the wind (windward) and negative pressure occurs overall on the side away from the wind (lee). If the vehicle is hit by a sudden gust, pressure distributions begin to build up gradually along the windward and lee sides of the vehicle [60]. Figure 16.28 illustrates this diagrammatically.

Immediately after a crosswind gust occurs, the pressure starts to increase along the windward side. After a short time, it stabilizes into a steady pressure distribution, applying a positive pressure overall to the windward side. The pressure change along the lee side does not begin until the wind gust has passed across the width of the vehicle. Negative pressures predominate along the lee side where, compared to the windward side, the lee-side pressure distribution starts to develop more slowly. The transition to a steady pressure distribution does not occur until a significantly later point in time. The difference in times, at which the transition to a steady pressure distribution on windward side and lee side has occurred, can amount to up to 0.5 s [60].

Figure 16.29 illustrates the effect of this time difference on the lateral force [51]. Pressures along the sides of a vehicle were measured under natural crosswind conditions. At the same time, the magnitude and direction of the incident air jet on the vehicle were recorded using a multi-hole probe mounted on the roof. Equation 16.9 was used to calculate the lateral wind-forces from the data measured by the multi-hole probe. These lateral wind-forces would arise if the pressure distributions along the windward and lee sides had developed simultaneously. On the other hand, lateral wind-forces were also derived from the measured pressures. These lateral wind-forces actually occur, that is, they involve the windward and lee side effects described earlier. A comparison of both lateral wind-forces (obtained from incident flow data and from measured pressures) indicates differences with regard to their characteristics as well as their magnitude.

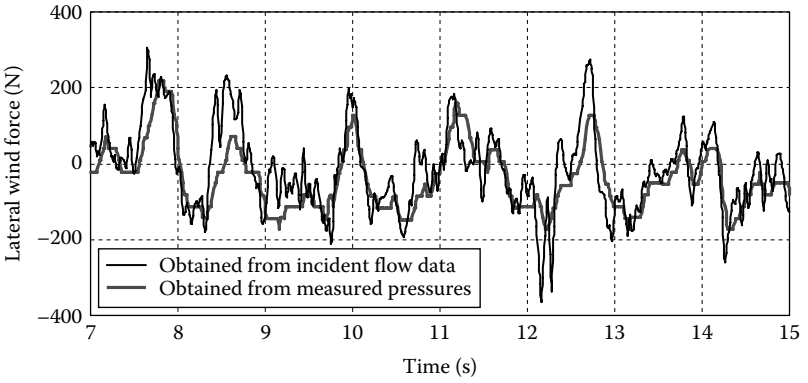


FIGURE 16.29 Derivation of stochastic lateral wind-forces by measuring pressures on the outer skin of the vehicle and measuring the incident airflow. (From Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.)

The dynamics of the pressure formation along the sides of a vehicle depend on the basic shape of the vehicle. This means that different body shapes can be excited differently by the same crosswind disturbance. The pressure formation in the presence of stochastic crosswind excitation is a central area of research into unsteady aerodynamics [58–61].

In experimental studies of the crosswind behavior, a measurement method can be used that acquires the dynamics of the pressure changes on the windward and lee sides. In simulations, however, the crosswind disturbance has to be modeled, that is, the excitation signal from the crosswind is artificially generated in the form of the product τv_{res}^2 (Section 16.5.1). This offers the benefit of being able to study not only a stochastic crosswind but also specific, selected crosswind phenomena. However, this does not take into consideration the dynamics of the crosswind excitation as described earlier.

If the dynamics of the crosswind excitation are ignored, it is inevitably "assumed" that, on the one hand, there is no time difference between excitations on the windward and lee sides and, on the other hand, different body shapes are excited in the same way. However, this is not the case in the real world. This situation poses a general problem when attempts are made to study crosswind disturbances, particularly of stochastic nature [51,61].

It is possible, in principle, to determine the dynamics of the excitation by computing the airflow and providing the simulation with these data. However, since a time-based phenomenon is involved, a transient airflow calculation has to be made. This must cover a sufficient time span, particularly where stochastic wind disturbances are concerned, so that it is possible to evaluate the results in the frequency domain. This generally results in considerable cost and effort, since it cannot be managed in an acceptable time frame, even with the most up-to-date computer capacity.

One promising possibility is to abandon artificially generated crosswind excitations and, instead, to use measured signals. This ensures, as a first step, that the simulation contains a realistic dynamic representation of the pressure formation on the windward and lee sides. Ideally, measurements can be carried out with the vehicle being tested under stochastic crosswind conditions, and the wind signal excitations recorded in this way can be fed into the simulation as input parameters. If this is not possible, for example, because a prototype is being studied, existing measurements from similar body shapes could be used [51].

16.7 CONCLUSION

Aerodynamics exerts a permanent effect on vehicle dynamics. Low drag reduces fuel consumption and increases performance. The aim of aerodynamic development of a vehicle is therefore to minimize the drag. Lift, pitch moment, and lateral components of the resulting wind-force are decisive parameters in lateral vehicle dynamics and therefore for the interaction between driver and vehicle. In optimizing them, note must be taken of the interactions between lift and drag.

The aerodynamic forces and moments depend on the basic shape of the vehicle and, in particular, on the shape of the rear end. Notchback vehicles have comparatively low drag and tend to have moderate lift, while fastbacks have both high drag and high lift particularly on the rear axle. Compared with notchbacks, squareback vehicles have a higher drag and a tendency to low lift or even negative lift.

One component of drag arises from the formation of longitudinal vortices that occur primarily at the rear end. This component is called vortex drag. It depends quadratically on the lift induced by the longitudinal vortices. Due to the dependency of drag and lift on each other, different rear end shapes must be treated differently for aerodynamic development. Measures to reduce drag applied to a vehicle with positive lift may have the opposite effect to a vehicle with negative lift.

Like all other aerodynamic forces and moments, lift also increases quadratically with road speed and exerts the wheel load. As a result, lateral control capability is reduced. Normally, the loading on front and rear axle is influenced by lift to a different extent. That influences the intrinsic steering behavior of the vehicle. Fluctuations in lift, for example, triggered by instationary flow conditions, have the same effect due to the regressive relationship between normal tire force and lateral tire control force.

In an oblique airstream, as, for example, crosswind, the vehicle experiences a lateral force, a yaw, and a roll moment. This causes the vehicle to veer off course and forces the driver to steer into it.

This process had been observed over a long period of time under quasi-steady conditions. During the last years, however, an instationary consideration of this phenomenon termed "crosswind sensitivity" has been emphasized. From a quasi-steady point of view, lateral force and yaw moment should be as low as possible. Within an instationary consideration, a generalization of this aspect cannot be made.

Thus, the entire system consisting of the driver and vehicle has to be studied. It is not sufficient simply to minimize the vehicle reactions to crosswind by aerodynamic and vehicle dynamics measures. Taking the system as a whole, the driver's influence can negate or reverse the measures taken to optimize the vehicle. Thus, the driver is drawn closer to the center of considerations. By evaluating his or her steering interventions, it is possible to investigate and optimize the interaction between driver and vehicle. A suitable driver model, which can adapt its steering characteristics to the driven vehicle, therefore, has to be incorporated into a virtual study.

A further aspect that should be taken into account is the unsteady character of the aerodynamic excitation of the vehicle. The time-dependent behavior of the pressure formation on the windward and lee sides is a complication in trying to accurately describe the vehicle's excitation. Furthermore, this behavior over time displays a dependency on the basic shape of the vehicle. This is therefore particularly important when different body shapes are being studied using virtual techniques.

This chapter demonstrates clearly that aerodynamic considerations, particularly with regard to unsteady lift and crosswind behavior, must be incorporated into the development process at the earliest possible stage of vehicle design. Furthermore, it is clear from the example of crosswind behavior that a more precise description is needed of the interface between aerodynamics and vehicle motion dynamics.

SYMBOLS

A	Frontal area (m^2)
a_Z	Lateral acceleration (m/s^2)
c_D	Aerodynamic drag coefficient
c_f	Lateral stiffness of front axle (N/rad)
c_L	Aerodynamic lift coefficient
c_{Lf}	Aerodynamic lift coefficient for front axle
c_{Lr}	Aerodynamic lift coefficient for rear axle
c_r	Lateral stiffness of rear axle (N/rad)
c_S	Aerodynamic side force coefficient
c_{Sf}	Aerodynamic side force coefficient for front axle
c_{Sr}	Aerodynamic side force coefficient for rear axle
c_{Mx}	Aerodynamic roll moment coefficient
c_{My}	Aerodynamic pitch moment coefficient
c_{Mz}	Aerodynamic yaw moment coefficient
$e_{CG,x}$	Lever arm of wind-force to *CG* in x direction (m)
F_f	Lateral force on front axle (N)
F_r	Lateral force on rear axle (N)
F_S	Lateral wind-force (N)
f	Frequency (Hz)
f_R	Resonant frequency (Hz)
$f_{R,opt}$	Optical resonant frequency of driver intervention (Hz)
G	Response function
G_F	Response function of the driver to crosswind excitation
G_{δ_L}	Response function of the vehicle to steering inputs
$G_{\tau v_{res}^2}$	Response function of the vehicle to crosswind excitation
g	Gravitational acceleration (m/s^2)
i_L	Steering ratio
J_z	Inertia around the vertical axis (kg m^2)

$j\omega$	Complex component of the frequency domain
K_{ges}	Evaluation criterion for the overall assessment
K_G	Evaluation criterion for the intensity of vehicle reactions
K_L	Evaluation criterion for the steering effort
K_P	Evaluation criterion for the driver's steering response
K_V	Evaluation criterion for the effectiveness of the driver's intervention
k_S	Gradient of the linearized aerodynamic side force coefficient
l_0	Wheelbase (m)
l_f	Distance between *CG* and front axle (m)
l_r	Distance between *CG* and rear axle (m)
M_x	Aerodynamic roll moment (Nm)
M_y	Aerodynamic pitch moment (Nm)
M_z	Aerodynamic yaw moment (Nm)
m	Vehicle mass (kg)
p	Pressure (bar)
r_{CG}	Track radius of the *CG* (m)
$T_{\dot{\psi},\max}$	Peak response time of yaw velocity (s)
t	Time (s)
v	Path speed of the *CG* (m/s)
v_f	Path speed of front axle (m/s)
v_r	Path speed of rear axle (m/s)
v_{res}	Resulting incident flow velocity (m/s)
v_{veh}	Flow velocity due to vehicle motion (m/s)
v_{Wind}	Wind speed (m/s)
α_f	Sideslip angle of front axle (composed of tire slip, elasto-kinematics, and steering elasticity) (rad)
α_r	Sideslip angle of rear axle (composed of tire slip and elasto-kinematics) (rad)
β	Slip angle (rad)
ϑ	Damping factor
δ	Damping time constant (1/s)
δ_L	Steering angle (rad)
ζ	Angular velocity (1/s)
ρ	Air density (kg/m^3)
τ	Resulting incident flow angle (rad)
τv_{res}^2	Factor to express stochastic vehicle excitation (radm^2/s^2)
τ_{Wind}	Wind angle (rad)
ψ	Yaw angle (rad)
$\dot{\psi}$	Yaw velocity (rad/s)
$\dot{\psi}_{\delta_L}$	Driver-induced yaw velocity (rad/s)
$\dot{\psi}_{\tau v_{res}^2}$	Wind-induced yaw velocity (rad/s)
ω	Angular frequency (rad/s)
ω_0	Undamped natural angular frequency (rad/s)
ω_d	Damped natural angular frequency (rad/s)
ω_R	Resonant angular frequency (rad/s)

REFERENCES

1. Hucho, W.-H., *Aerodynamik des Automobils*, 5th edn., Vieweg-Verlag, Wiesbaden, Germany, 2005.
2. Basara, B. and Tibaut, P., Time dependent versus steady state calculation of external aerodynamics, in *The Aerodynamics of Heavy Vehicles: Trucks, Buses and Trains*, R. McCallen, F. Browand, and J. Ross (eds.), Springer-Verlag, Berlin, Germany, 2004.

3. Hucho, W.-H., Aerodynamics of road vehicles—A challenge for computational fluid dynamics, *Proceedings of the 1st European Automotive CFD Conference*, Bingen, Germany, 2003.
4. Islam, M., Simulation of unsteady flows, in *Progress in Vehicle Aerodynamics*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Stuttgart, Germany, 2004.
5. Ahmed, S. R., Wake structure of typical automobile shapes, *Journal of Fluids Engineering "Transactions of the ASME"* 103, 162–169, 1981.
6. Cogotti, A., Wake surveys of different car-body shapes with coloured isopressure maps, SAE, Warrendale, PA, 1984, Paper 840299.
7. Howell, J., Shape and drag, in *Using Aerodynamics to Improve the Properties of Cars*, International Short Course on Vehicle Aerodynamics at FKFS, Stuttgart, Germany, 1998.
8. Schlichting, H. and Truckenbrodt, E., *Aerodynamik des Flugzeuges*, Springer-Verlag, Berlin, Germany, 1959.
9. Howell, J. and Le Good, G., Vortex drag for a simple car-like shape, *Proceedings of the 5th MIRA International Aerodynamics Conference*, Birmingham, U.K., 2004.
10. Howell, J. and Le Good, G., Vortex drag for a simple bluff body at incidence in ground proximity, SAE, Warrendale, PA, 2005, Paper 2005-01-0869.
11. Wagner, A., Wickern, G., and Zörner, Ch., Einfluss der Bodensimulation auf den induzierten Widerstand von Fahrzeugen, *Proceedings of the 6th International Symposium on Automotive and Engine Technology*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Stuttgart, Germany, 2005.
12. Wickern, G., Wagner, A., and Zörner, C., Induced drag of ground vehicles and its interaction with ground simulation, SAE, Warrendale, PA, 2005, Paper 2005-01-0872.
13. Cogotti, A., A parametric study on the ground effect of a simplified car model, SAE, Warrendale, PA, 1998, SP-1318.
14. Hucho, W.-H., Janssen, L. J., and Emmelmann, H. J., The optimization of body details—A method for reducing the aerodynamic drag of road vehicles, SAE, Warrendale, PA, 1976, Paper 760186.
15. Janssen, L. J. and Hucho, W.-H., The effect of various parameters on the aerodynamic drag of passenger cars, British Hydromechanical Association "Advances in Road Vehicle Aerodynamics", Cranfield, U.K., 1973.
16. Potthoff, J., Drag and lift of modern vehicles, *Proceedings of the 1st Symposium on Road Vehicle Aerodynamics*, The City University, London, U.K., 1969.
17. Zwicker, G., Wickern, K., and Pfadenhauer, M., Zum Einfluss von Rädern und Reifen auf den aerodynamischen Widerstand von Fahrzeugen, *Haus der Technik, Tagung "Aerodynamik des Kraftfahrzeugs"*, Essen, Germany, 1995.
18. Künster, R., Potthoff, J., and Essers, U., The aero-acoustic wind tunnel of Stuttgart University, SAE, Warrendale, PA, 1995, SP-1078.
19. Vagt, J.-D. and Wolff, B., Das neue Messzentrum für Aerodynamik—Zwei neue Windkanäle bei Porsche. Teil 1: ATZ 89, 121–129, Teil 2: ATZ 89, 183–189, 1987.
20. Waudby-Smith, P., Bender, T., and Vigneron, R., The GIE S2A full-scale aero-acoustic wind tunnel, SAE, Warrendale, PA, 2004, Paper 2004-01-0808.
21. Wickern, G. and Lindener, N., The Audi aeroacoustic wind tunnel: Final design and first operational experience, SAE, Warrendale, PA, 2000, SP-1524.
22. Buchheim, R., Unger, R., Carr, G., Cogotti, A., Kuhn, A., and Nilsson, L. U., Comparison tests between major European wind tunnels. SAE, Warrendale, PA, 1980, Paper 800410.
23. Le Good, G., Passenger car wind tunnel testing: The correlation of ¼ scale model to full-scale results, *Proceedings of the Conference on "Vehicle Aerodynamics"*, The Royal Aeronautical Society, Loughborough, U.K., 1994.
24. Cogotti, F., Widdeke, N., Kuthada, T., and Wiedemann, J., A new semi-empiric pad correction procedure, *Proceedings of the 5th Internationales Stuttgarter Symposium "Kraftfahrwesen und Verbrennungsmotoren"*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Renningen, Germany, 2003.
25. Mercker, E. and Wiedemann, J., On the correction of interference effects in open jet wind tunnels, SAE, Warrendale, PA, 1996, SP-1145.
26. Mercker, E., Wind tunnels, test-sections dimensions and corrections, in *Progress in Vehicle Aerodynamics—Advanced Experimental Techniques*, J. Wiedemann and W.-H. Hucho (eds.), Expert-Verlag, Renningen, Germany, 2000.
27. Mercker, E., Soja, H., and Wiedemann, J., Experimental investigation on the influence of various ground simulation techniques on a passenger car. *Proceedings of the Conference on Vehicle Aerodynamics*, The Royal Aeronautical Society, Loughborough, U.K., 1994.
28. Potthoff, J. and Wiedemann, J., Die Straßenfahrt-Simulation in den IVK-Windkanälen—Ausführung und erste Ergebnisse, *Proceedings of the 5th Internationales Stuttgarter Symposium für Kraftfahrwesen und Verbrennungsmotoren*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Renningen, Germany, 2003.

29. Wiedemann, J. and Potthoff, J., The new 5-belt road simulation system of the IVK wind tunnels—Design and first results, SAE, Warrendale, PA, 2003, Paper 2003-01-0429.
30. Cogotti, A., Ground effect simulation for full-scale cars in the pininfarina wind tunnel, SAE, Warrendale, PA, 1995, SP-1078.
31. Howell, J. and Hickmann, D., The influence of ground simulation on the aerodynamics of a simple car model, SAE, Warrendale, PA, 1997, SP-1232.
32. Wickern, G., The effect of moving ground on the aerodynamic drag of a production car, *Proceedings of the SATA Conference*, South Bend, IN, 1991.
33. Wickern, G., Wagner, A., and Zoerner, C., Cooling drag of ground vehicles and its interaction with ground simulation, SAE, Warrendale, PA, 2006, Paper 2006-01-0339.
34. Wiedemann, J., The influence of ground simulation and wheel rotation on drag optimization—Potential for reducing mileage, SAE, Warrendale, PA, 1996, Paper 960670.
35. Krajnovic, S. and Davidson, L., Large eddy simulation of the flow around a ground vehicle body, SAE, Warrendale, PA, 2004, SP-1874, pp. 5–14.
36. Basara, B., Plimon, A., Bachler, G., and Brandstätter, W., Calculation of flow around a car body with various turbulence models, *Proceedings of the 1st MIRA Vehicle Aerodynamics Conference*, Birmingham, U.K., 1996.
37. Vandromme, D., Overview of turbulence models for CFD in industry. Industrial computational fluid dynamics, Lecture Series 1995-03, Von Karman Institute for Fluid Dynamics, Brussels, Germany, 1995.
38. Breitling, T., Jehle, E., Reister, H., and Schwarz, V., Development and evaluation of a numerical simulation strategy designed to support the early stage of the aerodynamic development process, SAE, Warrendale, PA, 2002, SP-1667.
39. Lührmann, L., Chancen und Grenzen der virtuellen Aerodynamik. *Proceedings of the Conference: Haus der Technik "Fahrzeugerprobung—Von der Straße in den Rechner*", München, Germany, 2003.
40. Wagner, A., Application of CFD within the Audi development process, Vortrag Universität Girona, Spanien, Spain, 2006.
41. Watanabe, T., Present and future of high performance computing, *Proceedings of the NEC User Group Meeting*, Tokyo, Japan, 1995.
42. Mitschke, M., *Dynamik der Kraftfahrzeuge Band C, Fahrverhalten*, 2nd edn., Springer-Verlag, Berlin, Germany, 1990.
43. Reimpell, J., *Fahrwerktechnik: Fahrzeugmechanik*, Vogel-Verlag, Würzburg, Germany, 1983.
44. Wagner, A., Die Bewertung der Fahrer-Fahrzeug-Interaktion als Auslegungskriterium in der Fahrwerkentwicklung, *Proceedings of the Conference Haus der Technik "Fahrwerktechnik"*, München, Germany, 2003.
45. Wallentowitz, H., *Vertikal-/Querdynamik von Kraftfahrzeugen*, 2. Auflage, Aachen, Germany, 1997.
46. Zomotor, A., Fahrwerktechnik: Fahrverhalten, 2. Auflage, Vogel Verlag, Würzburg, Germany, 1991.
47. Schaible, S., Fahrzeugseitenwindempfindlichkeit unter natürlichen Bedingungen, Dissertation, TH Aachen, Aachen, Germany, 1998.
48. Wallentowitz, H., Fahrer—Fahrzeug—Seitenwind, Dissertation, TU Braunschweig, Braunschweig , Germany, 1978.
49. ISO 20119: 2002, Road vehicles—Test method for the quantification of on centre handling.
50. Föllinger, O., *Regelungstechnik*, 7th edn., Hüthig-Verlag, Heidelberg, Germany, 1992.
51. Wagner, A., Ein Verfahren zur Vorhersage und Bewertung der Fahrerreaktion bei Seitenwind, Dissertation, Universität Stuttgart, Stuttgart, Germany, 2003.
52. Bösch, P., Der fahrer als regler, Dissertation, TU Wien, Vienna, Austria, 1991.
53. Wagner, A. and Wiedemann, J., Crosswind behavior in the driver's perspective, SAE, Warrendale, PA, 2002, Paper 2002-01-0086.
54. Götz, H., Crosswind facilities and procedures, SAE Information Report 95-1109, Warrendale, PA, 1995.
55. ISO 12021-1: 1996, Road vehicles—Sensitivity to crosswind—Part 1: Open-loop test method using wind generator input.
56. Dettki, F., Eine Bewertungsmethode für den Geradeauslauf von Pkw unter Berücksichtigung von Seitenwind, *Proceedings of the 4th International Symposium Stuttgart "Automotive and Engine Technology"*, M. Bargende and J. Wiedemann (eds.), Expert-Verlag, Stuttgart, Germany, 2001.
57. Plöchl, M. and Lugner, P., Reaction of vehicle–driver system to unexpected external disturbances, EAEC, Wien, Vienna, Austria, 1999, Paper STA99C201.
58. Chadwick, A., Garry, K., and Howell, J., Transient aerodynamic characteristics of simple vehicle shape by the measurement of surface pressures, SAE, Warrendale, PA, 2000, Paper 2000-01-0876.
59. Nguyen, T. M., Saunders, J. W., and Watkins, S., The sideways dynamic force on passenger cars in turbulent winds, SAE, Warrendale, PA, 1997, Paper 970405.

60. Ryan, A. and Dominy, R.G., The transient aerodynamic characteristics of a passenger vehicle passing through a crosswind jet, ImechE, University of Durham, Durham, U.K., 1997, Paper C524/117/97.
61. Tran, T. V., Determining the wind forces and moments acting on vehicles by means of pressure sensors, SAE, Warrendale, PA, 1990, Paper 900313.
62. Watkins, S. and Saunders, J., A review of the wind conditions experienced by a moving vehicle, SAE, Warrendale, PA, 1998, Paper 981182.

Part III

Pneumatic Tires and Contact Wheel-Road/Off-Road

17 Tire as a Vehicle Component

Hans B. Pacejka

CONTENTS

17.1 Introduction and Definitions 563
17.2 Brush Model at Steady-State Slip 565
17.2.1 Theory of the General Case of Combined Slip 567
17.3 Similarity Method 573
17.3.1 Pure Slip Conditions 574
17.3.2 Combined Slip Conditions 579
17.4 Magic Formula Tire Model 581
17.4.1 Model Description 582
17.4.2 Full Set of Equations 590
17.5 First-Order Tire Lag 596
17.5.1 Model Development 596
17.5.2 Linear Model 596
17.5.3 Semi-Nonlinear Model 600
17.5.4 Fully Nonlinear Model 601
17.5.5 Gyroscopic Couple 603
17.5.6 Enhanced Nonlinear Transient Tire Model 604
17.6 Short-Wavelength Intermediate Frequency Tire Model 607
17.6.1 Contact Model Non-Steady-State Behavior 608
17.6.2 Residual Stiffnesses and Belt Dynamics 610
17.6.3 Dynamic Tire Response to Short Road Unevennesses 613
17.7 Tire Steady-State and Dynamic Test Facilities 616
17.A Appendix: Sign Conventions for Force and Moment and Wheel Slip 622
References 623
Additional Recommended Literature 624

17.1 INTRODUCTION AND DEFINITIONS

In each of the three perpendicular directions, the tire is capable to generate forces. They are needed to carry the vehicle's weight, to steer, and to brake or accelerate the vehicle. Apart from the forces, that is, the normal load F_z, the lateral force F_y, and the longitudinal driving or braking force F_x, the tire produces moments about the corresponding axes. The aligning torque M_z acting about the vertical axis is the more important one. In Figure 17.1 these forces and moment acting from road to tire have been indicated. Forces and moments are the result of tire deflections that arise due to the mutual interaction between the wheel and road. Obviously, the normal or radial deflection of the tire is a direct result of the vertical motion of the wheel axle and the vertical profile of the road surface. The horizontal deflections of the tire are connected with the horizontal and rotational motions of the wheel that may induce longitudinal and lateral slip of the wheel with respect to the ground.

To become more precise, it is useful to introduce the so-called slip point S. This imaginary point is attached to the wheel rim, and at the instant of observation, the point is located in the lowest position, that is, on the radius passing through the so-called contact center or point of intersection C.

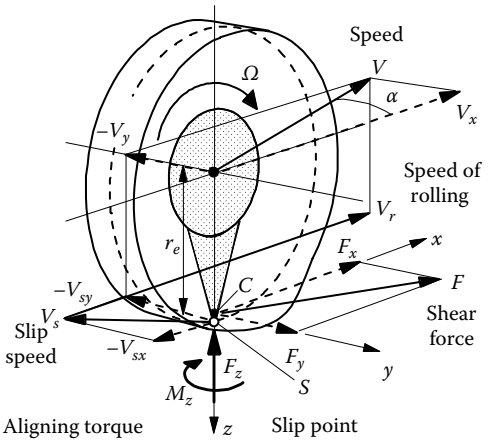


FIGURE 17.1 Tire forces and moment generated through tire deflections and wheel slip velocities.

This latter point forms the intersection of three planes: the road plane, the wheel center plane, and the plane passing through the wheelspin axis and perpendicular to the road plane. The slip point S is normally located slightly below the road surface. Its distance to the wheel center is defined as the slip radius r_s. Here, we take the slip radius equal to the so-called effective free rolling radius r_e. The horizontal velocity of the slip point is the slip speed of the wheel (rim) with respect to the road. Apparently, at free rolling (no driving or braking torque) the longitudinal component of the slip velocity equals zero. Then, the wheel speed of revolution $\Omega = \Omega_o$. When the wheel is being braked ($\Omega < \Omega_o$), point S moves in forward direction with longitudinal slip speed V_{sx}. When the slip point moves sideways, a lateral slip velocity V_{sy} arises. Note that the difference between lateral velocities of points S and C may be neglected. The resulting velocity V_s of point S represents the vector of the slip speed of the wheel. The slip speed components may be used directly to determine the resulting forces and moment. In practice, however, it appears in most cases to be more practical to use the components of the wheel slip vector as input quantities. These slip quantities are obtained by dividing the slip speed components by the forward speed V_x of the wheel center (or better of the contact center C):

$$\kappa = -\frac{V_{sx}}{V_x}\left(= \frac{r_e(\Omega - \Omega_o)}{V_x}\right) \tag{17.1}$$

$$\tan\alpha = -\frac{V_{sy}}{V_x} \tag{17.2}$$

Here, α is the slip angle as indicated in Figure 17.1. Other conventions may use different signs and symbols (cf. the Appendix). Depending on the type of testing equipment, one might define the longitudinal slip κ such that it equals zero at free rolling (as adopted here) or at vanishing longitudinal force $F_x = 0$. As a consequence, in the latter case a slightly shorter value of the slip radius results. The first definition is often used because of traditional testing procedures where the wheel can roll freely. The latter definition is used when the wheel rotational velocity is being controlled. It has the advantage that it can still be used even on very slippery surfaces where free rolling might not occur due to the low friction level not being able to overcome the tire rolling resistance.

The slip components α and κ together with spin slip φ due to wheel camber angle γ and possibly turn slip (path curvature $1/R$) may serve as input quantities in a tire model. With such a mathematical model, the resulting forces and moments can be calculated. The normal load and possibly the speed of travel are used as service parameters.

Tire models are used to calculate the tire forces and moments as responses to the wheel motion that may be expressed in terms of the various slip quantities. We may distinguish theoretical models based on physics of the tire construction and empirical models that are solely based on experimental findings. Also, combinations of both approaches are used in the development of the tire model.

The models that are in use nowadays range from very simple to highly complex descriptions of tire behavior. The model becomes more complex when we need to consider:

- Larger magnitudes of wheel slip, requiring a nonlinear description
- Combined slip as opposed to pure lateral or pure longitudinal slip
- Wheel camber and turn slip
- Wet or icy roads that make the friction coefficient dependency on speed no longer negligible
- Time-varying wheel motion that requires a non-steady-state/transient tire model
- High-frequency inputs and high speed of travel necessitating the inclusion of tire inertial properties
- Short-wavelength wheel oscillations and road irregularities requiring more detailed contact zone deflection considerations
- Simulations starting from or ending at standstill

In the subsequent sections, a brief description of modeling the tire as a vehicle component is given. Most of the material has been taken from the book "Tyre and Vehicle Dynamics" written by the author (cf. Ref. [1]). For further reading and for the study of vehicle dynamics applications, one is referred to that book and to the references listed at the end of this chapter. In the sequel, an overview will be given of some categories of tire modeling and tire dynamic behavior. We will first treat the relatively simple case of steady-state wheel slipping motions over a flat-level road surface by using the well-known physical brush model.

17.2 BRUSH MODEL AT STEADY-STATE SLIP

In the state-of-the-art paper by Pacejka and Sharp [2], a description is given of some recent tire models. The discussion was restricted to those models that were developed for use in vehicle dynamics studies at or close to steady-state conditions. To gain insight, let us consider a very simple physical model that can nicely explain how the forces and the moment are generated through the action of wheel slip in pure lateral or longitudinal directions or at a combination of these slip components (e.g., braking in a curve).

The model is the so-called brush tire model. The elastic bristles (similar to tread elements) are attached to the rim or to the assumedly rigid carcass that can only deflect vertically to accommodate the flat contact area. In the contact zone, the other ends of the bristles touch the road surface. Figure 17.2a depicts the deflected brush model. Three different cases of slip are considered. For each case the wheel slip speed vector (i.e., the speed of point S attached to the wheel rim) has been indicated. The base points of the elastic elements move parallel with respect to the road with that slip speed V_s (turn slip and camber not considered). As soon as the foremost element (at the right-hand side of the diagram) touches the road, the deformation starts to develop. The end point of the element that touches the surface remains in contact without sliding up to a point along the contact line where the frictional force that is maximally available is not enough to further deflect the element. At that transition point, the sliding zone is entered. In that zone, the element will continue to slide until it is lifted from the ground at the trailing edge. It may be noted that in the adhesion zone, where sliding does not occur, the contact line is straight and parallel to the speed vector of the wheel center (or better of the contact center C).

The model is simplified in that the friction coefficient is assumed to be constant (Coulomb friction) and that the stiffness of the elements is assumed to be the same in all horizontal directions. Further, a symmetric parabolic pressure distribution is considered. For the case of pure lateral slip,

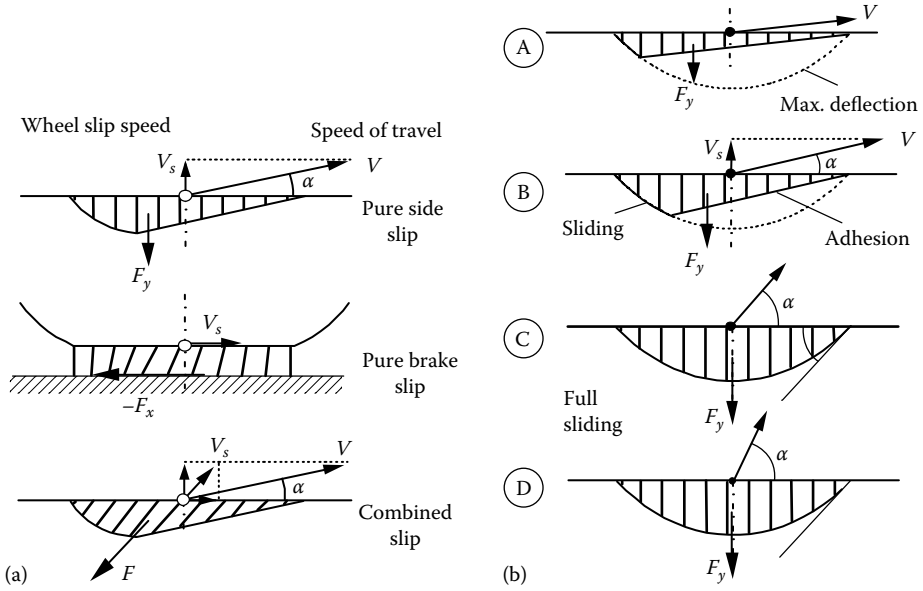


FIGURE 17.2 (a) Three modes of slip of the brush tire model. (b) Brush model at different levels of side slip.

Figure 17.2b shows how the deflection changes in magnitude and in distribution when the slip angle increases from very small to large. The initially approximately triangular shape (due to almost vanishing sliding) changes into a symmetric distribution when complete sliding is reached (case C).

We observe that the side force increases and that at the same time the pneumatic trail (the moment arm of the side force) decreases with increasing side slip. The resulting characteristics for the side force F_y and the aligning torque M_z are presented in Figure 17.3a. From the slip angle where complete sliding starts (case C), the side force is saturated and attains its maximum value equal to the friction coefficient μ times the vertical load F_z, while the moment becomes equal to zero. The curves appear to be very similar in nature to those established from measured data.

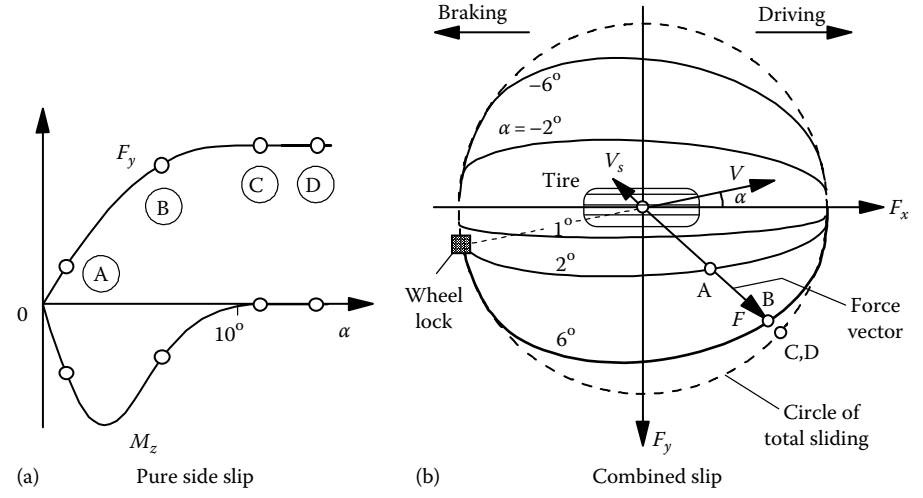


FIGURE 17.3 (a) The resulting side force and aligning torque characteristics of the model of Figure 17.2b with the corresponding cases A, B, C, and D indicated. (b) Combined slip characteristics for the side force versus fore and aft force at constant values of the slip angle. Cases A, B, C, and D of this isotropic tire model correspond in level of total slip with those indicated in Figure 17.2b.

The case of combined slip is analogous to the cases of pure lateral or longitudinal slip. Figure 17.3b shows the resulting F_y versus F_x diagram. Along the curves for constant slip angle α, the longitudinal slip κ varies. For a given ratio of tanα and κ, that is, for a given direction of the slip speed vector $\boldsymbol{V}_s$ (and the horizontal force vector $\boldsymbol{F}$ directed opposite to that slip speed), the four cases of increasing resulting slip A, B, C, and D have been indicated again.

17.2.1 Theory of the General Case of Combined Slip

To save space, we will not begin the theoretical analysis with one of the pure slip cases (cf. [1]), but start out with the general combined slip situation. For the analysis of the influence of longitudinal slip on the lateral force and moment generation, we will, for the sake of mathematical simplicity, restrict ourselves to the case of equal longitudinal and lateral stiffnesses of the tread elements (isotropic model), that is,

$$c_p = c_{px} = c_{py} \tag{17.3}$$

and equal and constant friction coefficients

$$\mu = \mu_x = \mu_y \tag{17.4}$$

Again a parabolic pressure distribution is considered.

Figure 17.4 depicts the deformations that arise when the tire model that runs at a given slip angle α is driven or braked. Due to the isotropic properties, the deflections are directed opposite to the slip

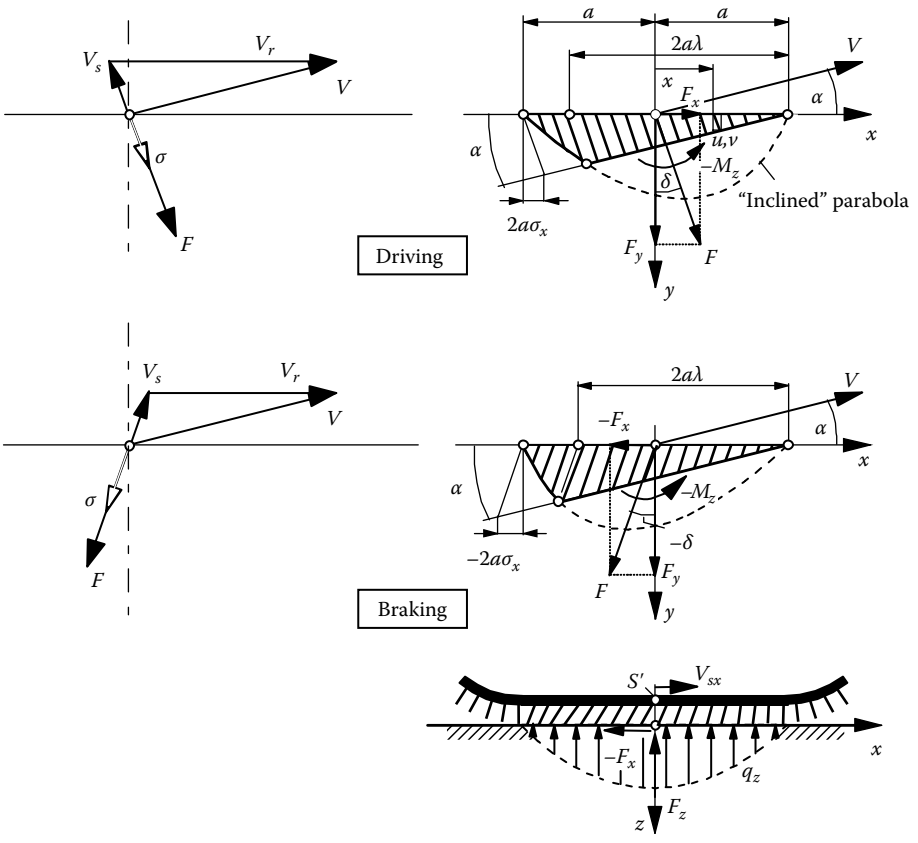


FIGURE 17.4 Vector diagram and deformation of the brush model running at a given slip angle for the cases of driving and braking.

speed vector $\boldsymbol{V}_s$, also in the sliding region. In this latter region, the tips of the elements slide over the road with sliding speed $\boldsymbol{V}_g$ directed opposite to the local friction force $\boldsymbol{q}$ (per unit contact length). The whole deformation history of a tread element while running through the contact area is a one-dimensional process along the direction of $\boldsymbol{V}_s$.

The velocity of progression of a base point through the contact length is the rolling speed V_r. The deflection rate of an element in the adhesion region is equal to the slip speed V_s. The time that elapses from the point of entrance to the point of a distance x in front of the contact center equals

$$\Delta t = \frac{a - x}{V_r} \tag{17.5}$$

with a denoting half the contact length. In this position, the deflection of an element that is still in adhesion becomes in vectorial form:

$$e = \begin{pmatrix} u \\ v \end{pmatrix} = \boldsymbol{V}_s \, \Delta t = -\frac{\boldsymbol{V}_s}{V_r}(a - x) \tag{17.6}$$

It seems natural at this stage to introduce the alternative (theoretical) slip quantity that reads in vectorial form:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_x \\ \sigma_y \end{pmatrix} = -\frac{\boldsymbol{V}_s}{V_r} = -\frac{1}{V_r}\begin{pmatrix} V_{sx} \\ V_{sy} \end{pmatrix} \tag{17.7}$$

with the linear speed of rolling

$$V_r = V_x - V_{sx} \tag{17.8}$$

The relations of these theoretical slip quantities with the practical slip quantities κ (= $- V_{sx}/V_x$) and $\tan\alpha$ (= $- V_{sy}/V_x$) are

$$\begin{aligned} \sigma_x &= \frac{\kappa}{1+\kappa} \\ \sigma_y &= \frac{\tan\alpha}{1+\kappa} \end{aligned} \tag{17.9}$$

The deflection of an element in the adhesion region now reads

$$e = (a - x)\,\boldsymbol{\sigma} \tag{17.10}$$

from which it is apparent that longitudinal and lateral deflections are governed by σ_x and σ_y, respectively, and independent of each other. This would not be the case if the deflection is expressed in terms of the practical quantities κ and α.

The local horizontal contact force acting on the tips of the elements (per unit contact length) reads

$$\boldsymbol{q} = c_p(a - x)\,\boldsymbol{\sigma} \text{ (adhesion region)} \tag{17.11}$$

As soon as

$$q = |\boldsymbol{q}| = \sqrt{q_x^2 + q_y^2} > \mu q_z \tag{17.12}$$

the sliding region is entered. Then the friction force vector becomes

$$\boldsymbol{q} = -\frac{\boldsymbol{V}_s}{V_s}\mu q_z = \frac{\boldsymbol{\sigma}}{\sigma}\mu q_z \text{ (sliding region)} \tag{17.13}$$

where

$$V_s = \sqrt{V_{sx}^2 + V_{sy}^2} \tag{17.14}$$

and

$$\sigma = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{17.15}$$

Similarly, the magnitude of the deflection of an element becomes

$$e = |\boldsymbol{e}| = \sqrt{u^2 + v^2} \tag{17.16}$$

The location of the point of transition from adhesion to sliding region at $x = x_t$ is obtained from the condition:

$$c_p e = \mu q_z \tag{17.17}$$

or

$$c_p \sigma (a - x_t) = \frac{3}{4}\mu F_z \frac{a^2 - x_t^2}{a^3} \tag{17.18}$$

which yields

$$x_t = \frac{4}{3}\frac{c_p a^3 \sigma}{\mu F_z} - a = a(2\theta\sigma - 1) \tag{17.19}$$

or with $\lambda = (a - x_t)/2a$ denoting the fraction of the contact length where sliding starts

$$\lambda = 1 - \theta\sigma \tag{17.20}$$

where the composite parameter θ reads

$$\theta = \frac{2}{3}\frac{c_p a^2}{\mu F_z} \tag{17.21}$$

From Equation 17.20 the slip σ_{sl} at which total sliding starts can be calculated. We get with $\lambda = 0$:

$$\sigma_{sl} = \frac{1}{\theta} \tag{17.22}$$

The magnitude of the total force $F = |\boldsymbol{F}|$ can now be obtained by integrating $c_p e$ over the contact length. We find

$$F = \mu F_z(1-\lambda^3) = \mu F_z\left\{3\theta\sigma - 3(\theta\sigma)^2 + \left(\theta\sigma\right)^3\right\} \quad \text{for } \sigma \le \sigma_{sl}$$
$$F = \mu F_z \quad \text{for } \sigma \ge \sigma_{sl} \tag{17.23}$$

that obviously follows the same course as the characteristic for F_y shown in Figure 17.3a. The force vector $\boldsymbol{F}$ acts in a direction opposite to $\boldsymbol{V}_s$ or $-\sigma$. Hence

$$\boldsymbol{F} = F\frac{\sigma}{\sigma} \tag{17.24}$$

from which the components F_x and F_y may be obtained.

The moment $-M_z$ is obtained by integration of $xc_p e_y$ over the contact length. If expressed in terms of multiplication of F_y with the pneumatic trail t, we have

$$M_z = -t(\sigma)\cdot F_y \tag{17.25}$$

The pneumatic trail t that indicates the distance behind the contact center C where the resultant side force F_y is acting appears to become

if $|\sigma| \le \sigma_{sl}$

$$t = -\frac{M_z}{F_y} = \frac{1}{3}a\frac{1-3\theta\sigma+3(\theta\sigma)^2-\theta\sigma^3}{1-\theta\sigma+(1/3)(\theta\sigma)^2} \tag{17.26}$$

and if $|\sigma| \ge \sigma_{sl}$

$$t = 0 \tag{17.27}$$

In Figures 17.5 and 17.6, the dramatic reduction of the pure slip forces (the side force and the longitudinal force, respectively) that occurs as a result of the simultaneous introduction of the other slip component (the longitudinal slip and the side slip, respectively) has been indicated. We observe a (almost) symmetric shape of these interaction curves. The peak of the side force versus

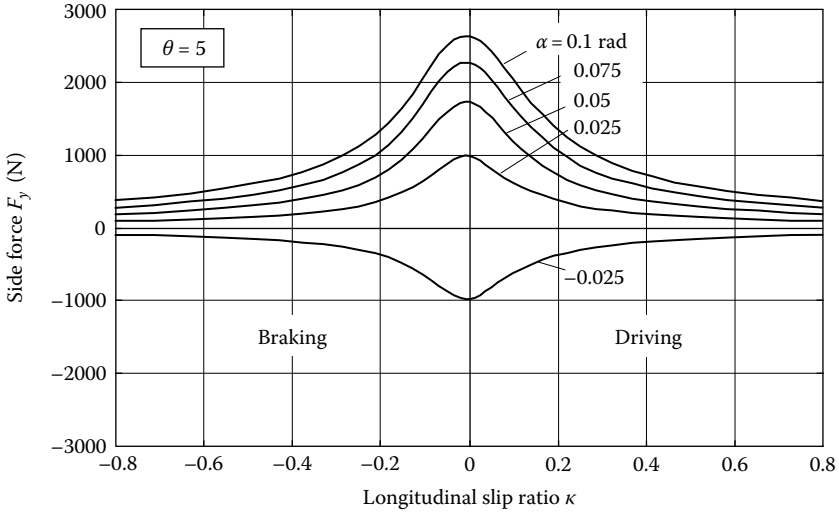


FIGURE 17.5 Reduction of side force due to the presence of longitudinal slip.

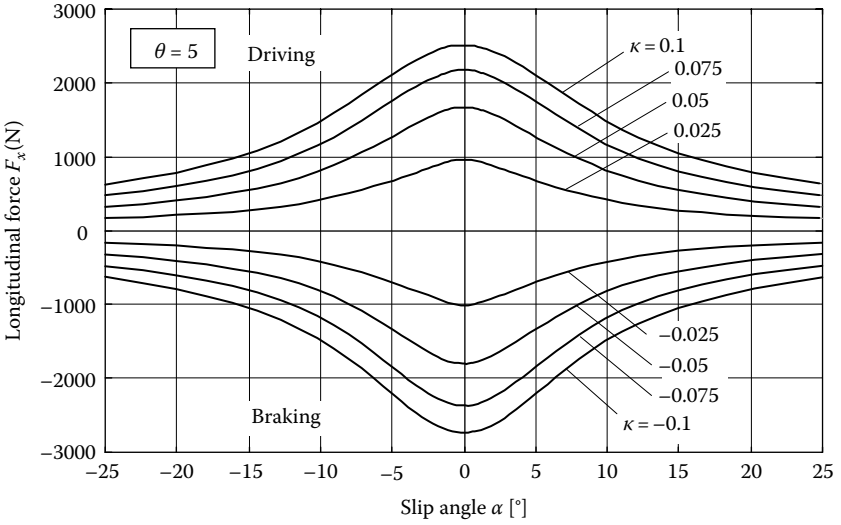


FIGURE 17.6 Reduction of longitudinal force due to the presence of side slip.

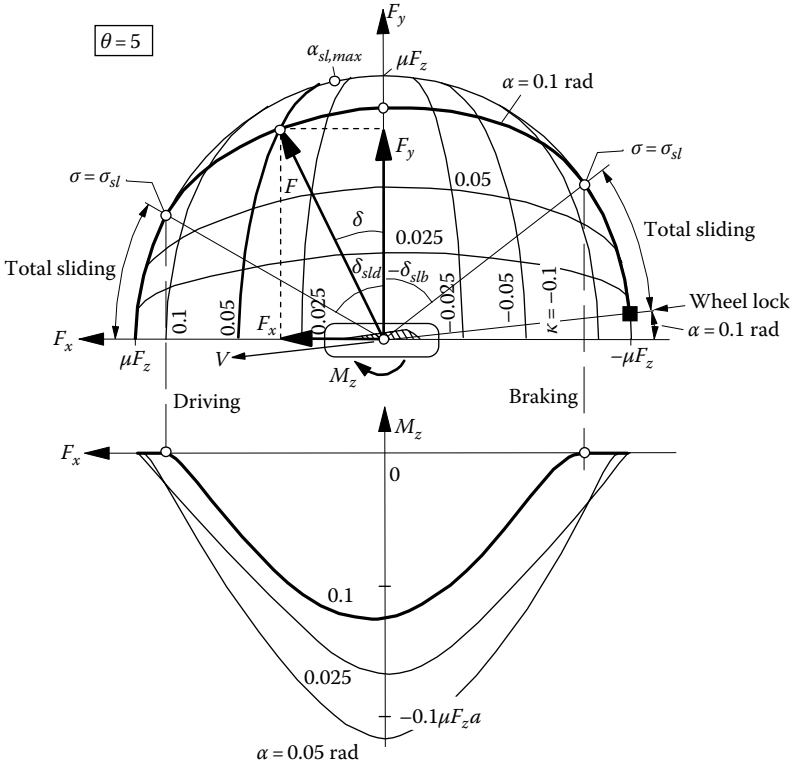


FIGURE 17.7 Cornering force and aligning torque as functions of longitudinal force at constant slip angle α or longitudinal slip κ.

longitudinal slip curves at constant values of the slip angle appears to be slightly shifted toward the braking side. This phenomenon will be further discussed in connection with the alternative representation of the same results according to Figure 17.7. At very large longitudinal slip, that is, when $|V_{sx}|/V_x \to \infty$, the side force approaches zero and the same occurs for the longitudinal force when the lateral slip tanα goes to infinity ($\alpha \to 90°$) at a given value of the longitudinal slip because

obviously in that case with $V_x \to 0$, also the longitudinal slip speed must vanish. For a wheel that is locked ($\Omega = 0$) with $V_{sx} = V_x$ and $\kappa = -1$, we have $F_y = \mu F_z \sin\alpha$ and $F_x = -\mu F_z \cos\alpha$.

In Figure 17.7, the calculated variations of F_y and M_z with F_x have been plotted for several fixed values of α. Also the curves for constant κ have been indicated. At braking, F_y appears to be a little larger than at driving. This would at least be true for those pairs of driving and braking conditions that show the same side slip angle and the same magnitude of the deviation angle δ of the slip velocity vector and thus of the force vector with respect to the y-axis. Then, the slip speeds V_s are equal in magnitude, but at braking the speed of rolling V_r is obviously smaller. When considering the definition of the theoretical slip σ (17.7) and the functional relationship (17.23) with the force F, it becomes clear that F must be larger in the case of braking because of the then larger magnitude of σ. Finally, the case of wheel lock is pointed at. Then, the force vector $\boldsymbol{F}$ that in magnitude is equal to μF_z is directed opposite to the speed vector of the wheel $\boldsymbol{V}$ that then coincides with the slip speed vector $\boldsymbol{V}_s$.

Experimental evidence supports the nature of the theoretical curves for the forces as shown in Figure 17.7. Often, the shape appears to be more asymmetric than predicted by the simple brush model. This may be due to a slight increase in the contact length while braking (making the tire stiffer) and by the brake force-induced slip angle of the contact patch at side slip. This is accomplished through the torsion of the carcass caused by the moment about the vertical axis that is exerted by the braking force that has shifted its line of action due to the lateral deflection that is associated with the side force.

The moment curves presented in the lower diagram of Figure 17.7 show a more or less symmetric bell shape. As expected, the aligning torque becomes equal to zero when total sliding occurs ($\sigma \geq \sigma_{sl}$). In cases observed in practice, these computed moment characteristics may appreciably deviate from the experimental ones. The model may be improved by introducing a lateral shift of the base line where the tread elements are attached. This shift corresponds to the lateral carcass deflection occurring at side slip under the action of the lateral force and may give rise to a sign change of the moment in the braking range of the characteristic (cf. [1]).

Although the force diagrams appear quite realistic, this simple brush model is in general not suitable to produce quantitatively correct and thus acceptable results. Considerable deviations may occur and the following important features that are not included in the simple brush model may be responsible for these (cf. Figure 17.8): unequal stiffnesses in x and y directions, nonconstant friction coefficient, lateral and bending compliance of the carcass and belt, nonsymmetric and nonconstant

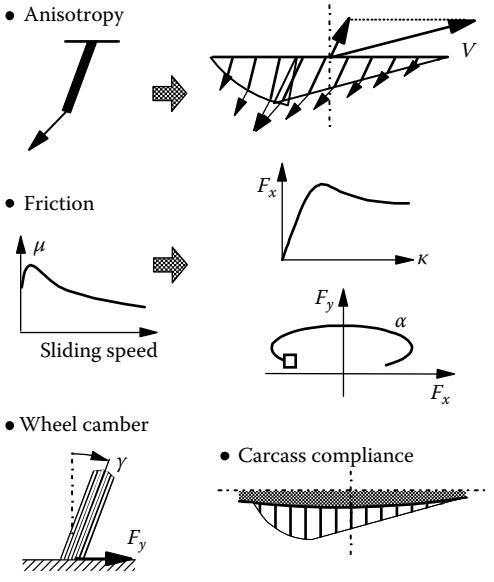


FIGURE 17.8 Physical factors that complicate modeling of the tire.

pressure distribution, and tread width. Introducing these factors in a tire model would considerably increase its complexity. Also, the introduction of wheel camber and turn slip would complicate the matter. To be able to handle these complicating features, computer models using discrete elements, the deflections of which are calculated during the simulation process, may be employed (cf. Pacejka et al. [1–3]). Some of these physical models may become highly advanced and can be used to simulate complex dynamic conditions (cf., e.g., Gipser [4,5] and Oertel [6,7]).

An alternative way to achieve a better mathematical description is the use of empirical models. For modeling steady-state conditions, two approaches will be discussed in the subsequent two sections: First, the similarity method and subsequently the more advanced Magic Formula tire model will be treated.

17.3 SIMILARITY METHOD

The method to be discussed in this section is based on the observation that the pure slip curves remain approximately similar in shape when the tire runs at conditions that are different from the reference condition. The reference condition is defined here as the state where the tire runs at its rated (nominal) load (F_{zo}), at camber equal to zero ($\gamma = 0$), at either free rolling ($\kappa = 0$) or side slip equal to zero ($\alpha = 0$), and on a given road surface (μ_o). A similar shape means that the characteristic that belongs to the reference condition is regained by vertical and horizontal multiplications and shifting of the curve. A demonstration that in practice similarity indeed approximately occurs is given by Radt and Milliken [8]. Figure 17.9 presents the results when the side force and the slip angle are normalized resulting in the nondimensional quantities shown along the axes. The raw data have been processed to make the characteristics pass through the origin of the graph. The curve results from a Magic Formula fit. The parameters B', C', D', and E' for the nondimensional side force (subscript y) have been used in the nondimensional version of the Magic Formula given further on. Also for the aligning torque, it can be shown that normalized curves appear to become very much the same.

The resulting model is through its simplicity relatively fast. It is capable to represent pure slip conditions rather well including the influence of a camber angle. The description of the situation at combined lateral and longitudinal slip is qualitatively satisfactory. Quantitatively, however, deviations may occur at higher levels of the combined two slip values.

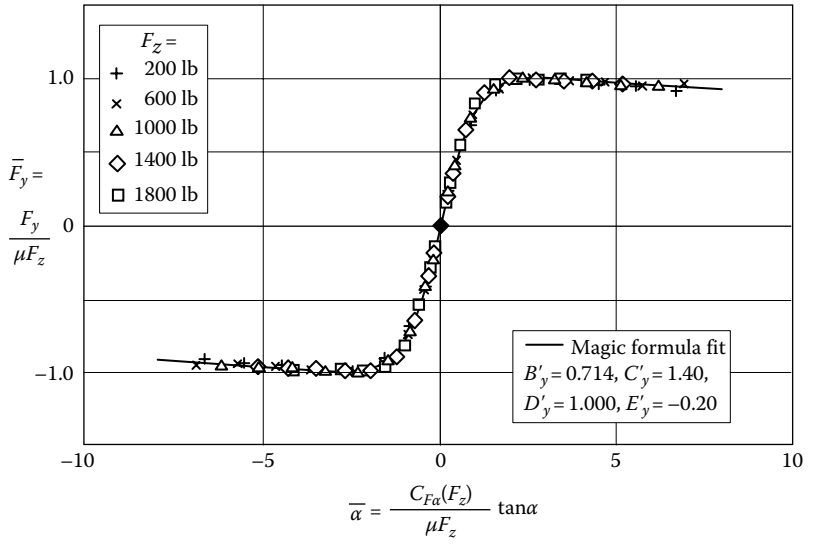


FIGURE 17.9 Result of Radt's nondimensionalization of tire characteristics showing that the tire side force characteristics measured at different loads reduce to virtually the same curve when the force and the slip angle are normalized as indicated.

17.3.1 Pure Slip Conditions

The functions representing the reference curves that are found at pure slip conditions are designated with the subscript *o*. We have, for instance, the reference function $F_y = F_{yo}(\alpha)$ that represents the side force versus slip angle relationship at nominal load F_{zo}, with longitudinal slip and camber equal to zero and the friction level represented by μ_o. We may now try to change the condition to a situation at a different wheel load F_z. Two basic changes will occur with the characteristic: (1) a change in level of the curve where saturation of the side force takes place (peak level) and (2) a change in slope at vanishing side slip ($\alpha = 0$). The first modification can be created by multiplying the characteristic both in vertical and in horizontal directions with the ratio F_z/F_{zo}. The horizontal multiplication is needed to not disturb the original slope. We obtain for the new function

$$F_y = \frac{F_z}{F_{zo}} F_{yo}(\alpha_{eq}) \tag{17.28}$$

with the equivalent slip angle:

$$\alpha_{eq} = \frac{F_{zo}}{F_z}\alpha \tag{17.29}$$

Because, obviously, the derivative of F_{yo} with respect to its argument α_{eq} at $\alpha_{eq} = 0$ is equal to the original cornering stiffness $C_{F\alpha o}$, we find for the derivative of F_y with respect to α the same value for the slope at $\alpha = 0$:

$$\frac{\partial F_y}{\partial \alpha} = \frac{F_z}{F_{zo}} \frac{dF_{yo}}{d\alpha_{eq}} \frac{\partial \alpha_{eq}}{\partial \alpha} = \frac{dF_{yo}}{d\alpha_{eq}} = C_{F\alpha o} \tag{17.30}$$

which proves that the slope at the origin of the characteristic is not affected by the successive multiplications. The second step in the manipulation of the original curve is the adaptation of the slope. This is accomplished by a horizontal multiplication of the newly obtained characteristic. This is done by multiplying the argument with the ratio of the new and the original cornering stiffness. Consequently, the new argument reads

$$\alpha_{eq} = \frac{C_{F\alpha}(F_z)}{C_{F\alpha o}} \frac{F_{zo}}{F_z}\alpha \tag{17.31}$$

Together with Equation 17.28 we have the new formulation for the side force versus slip angle relationship at the new load F_z. In Figure 17.10a the two steps taken to obtain the new curve have been illustrated. Here, the nominal load $F_{zo} = 3000$ N and the new load $F_z = 1500$ N.

In the same figure the characteristic of the aligning torque has been adapted to the new condition. For this, we use the knowledge gained when working with the brush model (cf. Figure 17.3a). More specifically, we will obey the rule that according to the theory, obviously, the point where the M_z curve reaches the α-axis lies below the peak of the F_y curve. According to Equation 17.22 this occurs at $\tan\alpha = 1/\theta$. This requirement means that the same equivalent slip angle (17.31) must be used as for the argument of the F_y function (17.28). Secondly, we will use information on the new value of the aligning stiffness $C_{M\alpha}(F_z)$. The reference curve for M_{zo} that is used is more realistic than the theoretical curve of Figure 17.3a. Typically, the moment changes its sign in the larger slip angle range approximately where F_y reaches its peak.

With the same equivalent slip angle and the new aligning stiffness, we obtain the expression for the new value of the aligning torque according to the similarity concept:

$$M_z = \frac{F_z}{F_{zo}} \frac{C_{M\alpha}(F_z)}{C_{M\alpha o}} \frac{C_{F\alpha o}}{C_{F\alpha}(F_z)} M_{zo}(\alpha_{eq}) \tag{17.32}$$

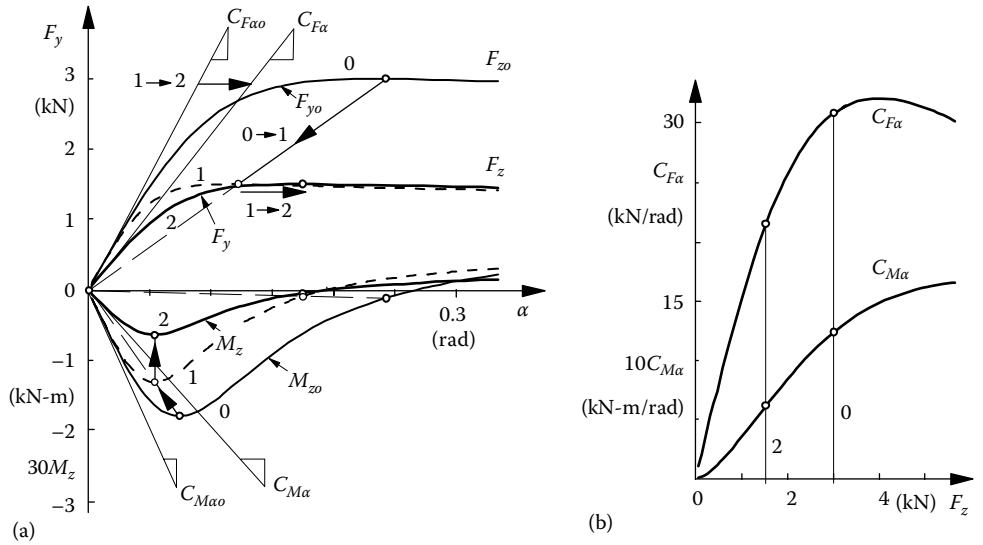


FIGURE 17.10 (a) Using the similarity method to adapt F_y and M_z curves to a new level of load. (b) Cornering and aligning stiffness versus wheel load.

The first and third factors being the inverse of the multiplication factor used in (17.31) are needed to maintain the original slope. The second factor multiplies the intermediate M_z curve (1 in Figure 17.10a) in vertical direction to adapt the slope to its new value (2) while not disturbing the α scale. It may be noted that the combined second and third factor equals the ratio of the new and the original values of the pneumatic trail t/t_o.

For the calculations connected with Figure 17.10a, the reference characteristics for F_{yo}, M_{zo}, and $C_{F\alpha}$ have been described by means of the Magic Formula-type functions. For this occasion, the aligning stiffness is modeled as the product of a certain fraction of the contact length $2a$ and the cornering stiffness. The resulting characteristics for these four quantities as shown in Figure 17.10a and b are realistic. The following formulae have been used:

The side force at nominal load F_{zo}

$$F_{yo} = D_{yo} \sin[C_y \arctan \{B_{yo}\alpha - E_y(B_{yo}\,\alpha - \arctan(B_{yo}\alpha))\}] \tag{17.33}$$

with stiffness factor

$$B_{yo} = \frac{C_{F\alpha o}}{C_y D_{yo}} \tag{17.34}$$

the peak factor for the side force that, in general, is different from the one for the longitudinal force (as a consequence, we introduce μ_{yo} beside μ_{xo} in the peak factor)

$$D_{yo} = \mu_{yo} F_{zo} \tag{17.35}$$

and the cornering stiffness as function of wheel load F_z

$$C_{F\alpha} = c_1 c_2 F_{zo} \sin\left\{2 \arctan\left(\frac{F_z}{c_2 F_{zo}}\right)\right\} \tag{17.36}$$

The aligning torque at nominal wheel load

$$M_{zo} = D_{zo} \sin[C_z \arctan \{B_{zo}\alpha - E_z(B_{zo}\alpha - \arctan(B_{zo}\alpha))\}] \tag{17.37}$$

with stiffness factor

$$B_{zo} = \frac{-C_{M\alpha o}}{C_z D_{zo}} \tag{17.38}$$

the peak factor (a_o representing half the contact length at nominal load)

$$D_{zo} = c_3 a_o D_{yo} \tag{17.39}$$

and the aligning stiffness as function of wheel load F_z

$$C_{M\alpha} = tC_{F\alpha} = c_4 a C_{F\alpha} \tag{17.40}$$

where apparently the pneumatic trail $t = c_4 a$. Half the contact length a is assumed to be proportional with the square root of the wheel load:

$$a = a_o \sqrt{\frac{F_z}{F_{zo}}} \tag{17.41}$$

The resulting approximation of the pneumatic trail turns out to be quite adequate and might be considered to be used in the set of Magic Formula tire model in Equation 17.128 to be dealt with later on.

The values of parameters used for the calculations to be conducted have been listed in Table 17.1.

The next two items we will deal with are a change in friction coefficient from μ_{yo} to μ_y and the introduction of a camber angle γ. The first change can be handled by multiplying the curves in radial direction with factor μ_y/μ_{yo}. Together with the change in load, we find

$$F_y = \frac{\mu_y F_z}{\mu_{yo} F_{zo}} F_{yo}(\alpha_{eq}) \tag{17.42}$$

with

$$\alpha_{eq} = \frac{C_{F\alpha}(F_z)}{C_{F\alpha o}} \frac{\mu_{yo} F_{zo}}{\mu_y F_z} \alpha \tag{17.43}$$

The friction coefficient may be assumed to depend on (decay with) the slip velocity V_s. To model the situation on wet roads, a decaying function $\mu(V_s)$ may be employed (cf., e.g., ([1], Equations 17.110 and 17.95)).

TABLE 17.1
Parameter Values

F_{zo}	3000 N	C_y	1.3	C_x	1.5	c_1	8	c_5	1
a_o	0.08 m	E_y	−1	E_x	−1	c_2	1.33	c_6	0.3
b	0.07 m	C_z	2.3	μ_{yo}	1	c_3	0.25	c_7	100
r_e	0.30 m	E_z	−2	μ_{xo}	1.26	c_4	0.5	c_8	15

When considering experimental $F_y - \alpha$ characteristics at small camber angles, a reasonable result is expected when the F_y curve is shifted horizontally. Then, the peak side force remains unchanged, which is a reasonable assumption but not always supported by experimental evidence (where often but not always a slight increase in maximum side force is manifested) or by computations with a physical model with constant friction and finite contact width that show a slight decrease. For small angles, the camber thrust is approximated by the product of the camber stiffness and the camber angle: $F_{y\gamma} = C_{F\gamma}\gamma$. Consequently, the α shift should amount to

$$S_{Hy} = \frac{C_{F\gamma}(F_z)}{C_{F\alpha}(F_z)}\gamma \tag{17.44}$$

so that

$$\alpha^* = \alpha + S_{Hy} \tag{17.45}$$

which gives rise to α_{eq} (17.43) with α replaced by α^*.

For the representation of the aligning moment at the new conditions, the situation is more complex. From test results and more advanced calculations, it is observed that for not too large slip angles, the original curve M_{zo} tends to shift both sideways and upward. If the same equivalent slip angle and thus the same horizontal shift (17.44) as for the force are employed, the moment would become zero where also the side force curve crosses the α-axis. A moment equal to $-C_{M\alpha}\, S_{Hy}$ would arise at $\alpha = 0$. However, the moment should become equal to $C_{M\gamma}\gamma$, with a positive camber moment stiffness $C_{M\gamma}$. This implies that an additional vertical shift is required equal to

$$S_{Vz} = C_{M\gamma}(F_z)\gamma + C_{M\alpha}(F_z)S_{Hy} \tag{17.46}$$

This additional moment corresponds to the so-called residual torque M_{zr}, which is the moment that remains when the side force becomes equal to zero. For larger values of the slip angle, the additional moment should tend to zero. This may be easily accomplished by dividing (17.46) by a term like $1 + c_7\alpha^2$.

The dependencies of the camber stiffnesses on the vertical load may be assumed linear. We have

$$C_{F\gamma} = c_5 F_z, \quad C_{M\gamma} = c_6 \frac{b^2}{r_e} C_{F\kappa} \tag{17.47}$$

where

b denotes half the (assumedly constant) width of the contact patch

r_e is the effective rolling radius

$C_{F\kappa}$ is the longitudinal slip stiffness, which is considered to be linearly increasing with load

$$C_{F\kappa} = c_8 F_z \tag{17.48}$$

The camber moment stiffness $C_{M\gamma}$ mainly results from left and right opposite longitudinal slip in the contact area. In Figure 17.11, the different steps leading to the condition with a (small) camber angle have been demonstrated. First, the original curve (0) is moved to the left over a distance equal to the horizontal shift (17.44) yielding the intermediate curve (1). Then the residual torque is added: vertical shift (17.46) possibly made a function of slip angle by multiplying the shift with a decaying weighting function as suggested earlier. The load F_z and the friction coefficient μ_y have been kept equal to their reference values 3000 N and 1, respectively.

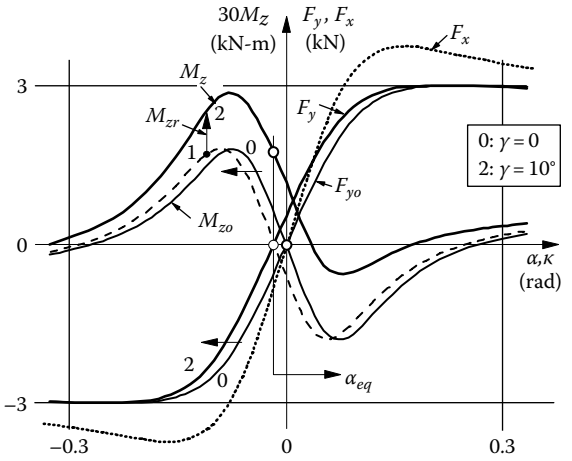


FIGURE 17.11 Introduction of a camber angle in the side force and the aligning torque pure slip characteristics according to the similarity method. The longitudinal force pure slip characteristic. ($F_z = F_{zo} = 3000$ N, $\mu_y = \mu_{yo} = 1$, $\mu_x = \mu_{xo} = 1.26$).

The longitudinal force at pure longitudinal slip may be modeled in the same way as we did for the side force. Equations similar to (17.33) through (17.35) hold for the reference function. An influence of camber on the longitudinal force, however, is assumed not to occur. In the same Figure 17.11, the characteristic for F_x versus κ has been drawn. The values of the relevant additional parameters for the longitudinal force and the moment have been listed in Table 17.1.

We finally have the following similarity formulae for pure slip conditions (either at longitudinal or at lateral slip) for the longitudinal force:

$$F_x = \frac{\mu_x F_z}{\mu_{xo} F_{zo}} F_{xo}(\kappa_{eq}) \tag{17.49}$$

with the equivalent longitudinal slip

$$\kappa_{eq} = \frac{C_{F\kappa}(F_z)}{C_{F\kappa o}} \frac{\mu_{xo} F_{zo}}{\mu_x F_z} \kappa \tag{17.50}$$

and for the side force

$$F_y = \frac{\mu_y F_z}{\mu_{yo} F_{zo}} F_{yo}(\alpha_{eq}) \tag{17.51}$$

with the equivalent slip angle, containing the horizontal shift (17.44),

$$\alpha_{eq} = \frac{C_{F\alpha}(F_z)}{C_{F\alpha o}} \frac{\mu_{yo} F_{zo}}{\mu_y F_z} \left(\alpha + \frac{C_{F\gamma}(F_z)}{C_{F\alpha}(F_z)} \gamma \right) \tag{17.52}$$

and for the aligning torque

$$M_z = \frac{\mu_y F_z}{\mu_{yo} F_{zo}} \frac{C_{M\alpha}(F_z)}{C_{M\alpha o}} \frac{C_{F\alpha o}}{C_{F\alpha}(F_z)} M_{zo}(\alpha_{eq}) + M_{zr} \tag{17.53}$$

with the residual torque corresponding to the vertical shift (17.46) provided with the reduction factor

$$M_{zr} = \frac{C_{M\gamma}(F_z) + t(F_z)C_{F\gamma}(F_z)}{1 + c_7\alpha^2}\gamma \tag{17.54}$$

17.3.2 Combined Slip Conditions

We will now address the problem of describing the situation at combined slip. The analysis of the brush model has gained considerable insight into the mechanisms that play a role and we will use the theoretical slip quantities $\sigma_{x,y}$ (17.9) and the magnitude σ (17.15) and we will adopt the concept similar to that of assessing the components of the resulting horizontal force according to (17.24) and use the pneumatic trail t as in (17.25). We have for the theoretical slip quantities with the α shift due to camber included

$$\sigma_x = \frac{\kappa}{1+\kappa} \tag{17.55}$$

$$\sigma_y^* = \frac{\tan\alpha^*}{1+\kappa} \tag{17.56}$$

and

$$\sigma^* = \sqrt{\sigma_x^2 + \sigma_y^{*2}} \tag{17.57}$$

with

$$\sigma^* = \alpha + \frac{C_{F\gamma}(F_z)}{C_{F\alpha}(F_z)}\gamma \tag{17.58}$$

In the ensuing theory, we will make use of the theoretical slip quantities (17.55, 17.56), which does produce the slope in the constant α curves in the F_y versus F_x diagram at $F_x = 0$. It should be noted, however, that experience has shown that the use of practical slip quantities s_x and s_y^* (i.e., κ and $\tan\alpha^*$ that may be obtained again by omitting the denominators in (17.55, 17.56) may already give very good results (cf. [1]). Note that when approaching wheel lock, $\sigma_x \to \infty$. This calls for an artificial limitation by adding a small positive quantity to the denominator of expression (17.55).

Because we are dealing with an in general non-isotropic tire, we have pure longitudinal and lateral slip characteristics that are not identical (cf. Figure 17.11). Still, we will take the lateral and longitudinal components but then of the respective pure slip characteristics F_{yo} and F_{xo}. If we want to use the theoretical slip quantities σ, we must have available pure slip characteristics with σ_x and σ_y as abscissa.

Next, we should realize that through (17.45) camber has been accounted for and that, as a consequence, σ_y is to be replaced by σ_y^* defined by Equation 17.56. The resulting force and moment functions are denoted by $F_{xo}(\sigma_x)$, $F_{yo}(\sigma_y^*)$, and $M_{zo}(\sigma_y^*)$. As illustrated in Figure 17.12, the "components" now become with σ^* denoting the magnitude of the theoretical slip vector (17.57):

$$F_x = \frac{\sigma_x}{\sigma^*}F_{xo}(\sigma^*), \quad F_y = \frac{\sigma_y^*}{\sigma^*}F_{yo}(\sigma^*) \tag{17.59}$$

The figure shows how the force vector arises from the individual pure slip characteristics. At small slip, in general, the vector does not run opposite with respect to the slip speed vector $\boldsymbol{V}_s$.

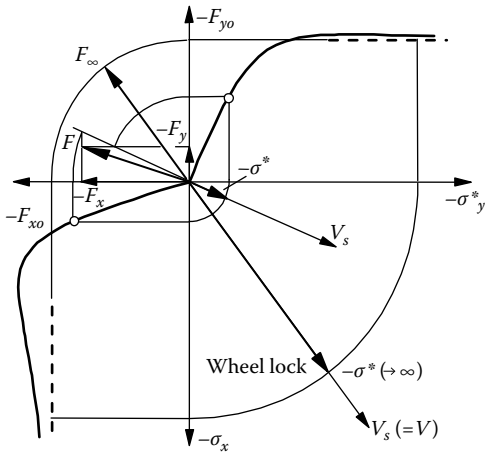


FIGURE 17.12 Construction of the resulting combined slip force from the pure slip characteristics according to Equation 17.59.

At wheel lock, however, the force vector does run opposite with respect to the slip speed vector because it is assumed here that the levels of the individual characteristics (asymptotes) are the same for the slip approaching infinity (in the Magic Formula we should have $\mu_{yo}\sin(\tfrac{1}{2}\pi C_y) = \mu_{xo}\sin(\tfrac{1}{2}\pi C_x)$). At large slip the theory apparently fails to properly account for the contribution of the camber angle. At wheel lock one would expect the side force to become zero at vanishing slip angle. This will not exactly be the case in the model due to the equivalent theoretical side slip defined according to Equation 17.56 with α^* given by (17.45).

For the aligning torque, the formulae have not been reproduced. We may refer to the original text [1].

With the similarity expressions for the pure slip forces $F_{x,yo}$, we finally obtain the following formulae for combined slip conditions. For the longitudinal force

$$F_x = \frac{\sigma_x}{\sigma^*}\frac{\mu_x F_z}{\mu_{xo} F_{zo}} F_{xo}(\sigma^x_{eq}) \tag{17.60}$$

with the equivalent slip

$$\sigma^x_{eq} = \frac{C_{F\kappa}(F_z)}{C_{F\kappa o}}\frac{\mu_{xo} F_{zo}}{\mu_x F_z}\sigma^* \tag{17.61}$$

and for the side force

$$F_y = \frac{\sigma^*_y}{\sigma^*}\frac{\mu_y F_z}{\mu_{yo} F_{zo}} F_{yo}(\sigma^y_{eq}) \tag{17.62}$$

with the equivalent slip

$$\sigma^y_{eq} = \frac{C_{F\alpha}(F_z)}{C_{F\alpha o}}\frac{\mu_{yo} F_{zo}}{\mu_y F_z}\sigma^* \tag{17.63}$$

For the case of a load different from the reference value and the presence of a camber angle, the combined slip characteristics for the tire side force have been assessed by using the preceding equations.

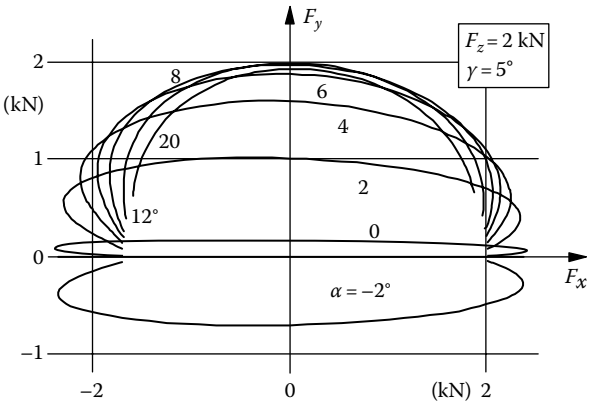


FIGURE 17.13 Cornering force versus longitudinal force at a new load and camber angle as computed with the similarity method.

The resulting diagrams with slip angle α used as parameter have been presented in Figure 17.13. Due to the use of the theoretical slip quantities (17.55, 17.56), we find that the curves of F_y show the slight slope near $F_x = 0$ also observed to occur with the physical brush model and sometimes more pronounced in experimentally assessed characteristics. The additional slope that arises when a camber angle is introduced for which the torsional compliance is responsible does not appear. A special formulation would be needed to represent this effect. In the next section, we will take care of this in connection with the Magic Formula tire model.

It may be further remarked that the similarity model given by Equations 17.60 through 17.63 together with (17.55) and (17.56) and the reference pure slip functions does represent the actual tire steady-state behavior correctly in the following extreme cases: (1) in pure slip situations, (2) in the linearized combined slip situation (small α and κ), and (3) in the case of a locked wheel ($\gamma = 0$). As illustrated in Figure 17.12, the model correctly shows that when the wheel is locked, the resulting force acts in a direction opposite to the slip speed vector $\boldsymbol{V}_s$ if in the original reference pure slip curves, the level of F_{xo} and of F_{yo} tends to the same level when both slip components approach infinity (governed by parameters μ and C in the Magic Formula). Other combinations of slip may give rise to deviations with respect to measured characteristics. It is noted that the approximate combined slip performance of the side and fore and aft forces can be represented without the necessity to have combined slip measurement data. In [1] the computed results have been compared with experimental data showing good agreement.

17.4 MAGIC FORMULA TIRE MODEL

A widely used semiempirical tire model to calculate steady-state tire force and moment characteristics for use in vehicle dynamics studies is based on the so-called Magic Formula. The development of the model was started in the mid-1980s. In a cooperative effort, TU Delft and Volvo developed several versions (Bakker et al. [9], Pacejka et al. [10]). In these models the combined slip situation was modeled from a physical view point. In 1993 Michelin (cf. Bayle et al. [11]) introduced a purely empirical method using Magic Formula-based functions to describe the tire horizontal force generation at combined slip. This approach was adopted in Delft Tire [12]. In the newer versions of "Delft Tire," the original description of the aligning torque is altered to accommodate a relatively simple physically based combined slip extension. The pneumatic trail is introduced as a basis to calculate this moment about the vertical axis (cf. Pacejka [12]). A complete listing of the model is given in the sequel.

17.4.1 Model Description

The general form of the formula that holds for given values of vertical load and camber angle reads

$$y = D\sin[C\arctan\{Bx - E(Bx - \arctan Bx)\}] \tag{17.64}$$

with

$$Y(X) = y(x) + S_V \tag{17.65}$$

$$x = X + S_H \tag{17.66}$$

where

Y is the output variable F_x, F_y, or possibly M_z
X is the input variable $\tan\alpha$ or κ
B is the stiffness factor
C is the shape factor
D is the peak value
E is the curvature factor
S_H is the horizontal shift
S_V is the vertical shift

The Magic Formula $y(x)$ typically produces a curve that passes through the origin $x = y = 0$, reaches a maximum, and subsequently tends to a horizontal asymptote. For given values of the coefficients B, C, D, and E, the curve shows an antisymmetric shape with respect to the origin. To allow the curve to have an offset with respect to the origin, two shifts S_H and S_V have been introduced. A new set of coordinates $Y(X)$ arises as shown in Figure 17.14. The formula is capable of producing characteristics that closely match measured curves for the side force F_y (and if desired also for the aligning torque M_z, cf. Equation 17.37) and for the fore and aft force F_x as functions of their respective slip quantities: the slip angle α and the longitudinal slip κ with the effect of load F_z and camber angle γ included in the parameters.

Figure 17.14 illustrates the meaning of some of the factors by means of a typical side force characteristic. Obviously, coefficient D represents the peak value (with respect to the central x-axis and for $C \geq 1$) and the product BCD corresponds to the slope at the origin ($x = y = 0$). The shape factor C controls the limits of the range of the sine function appearing in formula (17.64) and thereby determines the shape of the resulting curve. The factor B is left to determine the slope at the origin and is called the stiffness factor. The factor E is introduced to control the curvature at the peak and at the same time the horizontal position of the peak.

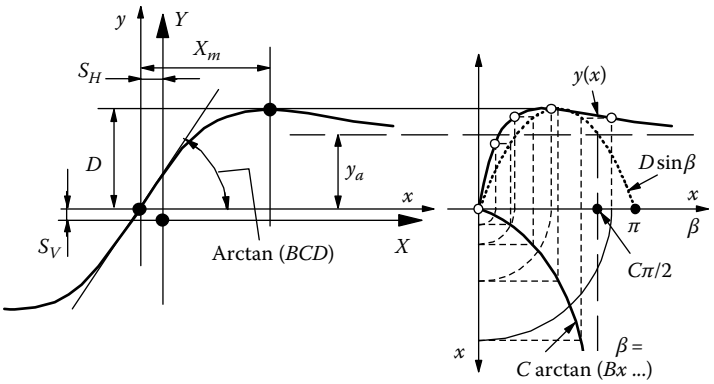


FIGURE 17.14 Curve produced by the original sine version of the Magic Formula, Equation 17.136. The meaning of curve parameters has been indicated.

From the heights of the peak and of the horizontal asymptote, the shape factor C may be computed:

$$C = 1 \pm \left(1 - \frac{2}{\pi} \arcsin \frac{y_a}{D} \right) \tag{17.67}$$

From B and C and location x_m of the peak, the value of E may be assessed:

$$E = \frac{Bx_m - \tan\{\pi/(2C)\}}{Bx_m - \arctan(Bx_m)} \quad (\text{if } C > 1) \tag{17.68}$$

The offsets S_H and S_V appear to occur when ply-steer and conicity effects and possibly the rolling resistance cause the F_y and F_x curves not to pass through the origin. Wheel camber may give rise to a considerable offset of the F_y versus α curves. Such a shift may be accompanied by a significant deviation from the pure antisymmetric shape of the original curve. To accommodate such an asymmetry, the curvature factor E is made dependent of the sign of the abscissa (x):

$$E = E_o + \Delta E \cdot \operatorname{sgn}(x) \tag{17.69}$$

Also the difference in shape that is expected to occur in the F_x versus κ characteristic between the driving and braking ranges can be taken care of. In Figure 17.15, the influence of the two shape factors C and E on the appearance of the curves has been demonstrated. The diagrams have been normalized by dividing y by D and multiplying x with CB making the curve peak level and initial slope independent of the parameters.

In rather extreme cases, the sharpness that can be reached by means of the function given by Equation 17.64 may not be sufficient. It turns out to be possible to considerably increase the sharpness of the curves by introducing an extra term in the argument of the arctan function. For this we may refer to the original text [1].

The various factors are functions of normal load and wheel camber angle. Several parameters appear in these functions. A suitable regression technique is used to determine their values from measured data according to a quadratic algorithm for the best fit (cf. Oosten and Bakker [13]). One of the important functional relationships is the variation of the cornering stiffness (almost exactly given by the product of the coefficients B_y, C_y, and D_y of the side force function: $BCD_y = K_{y\alpha} = C_{F\alpha} = \partial F_y/\partial\alpha$ at $\tan\alpha = -S_H$) with F_z and γ.

$$BCD_y = p_1 \sin \frac{2 \arctan(F_z/p_2)}{1 + p_3\gamma^2} \tag{17.70}$$

For zero camber, the cornering stiffness attains its maximum p_1 at $F_z = P_2$. In Figure 17.16, the basic relationship has been depicted. Apparently, for a cambered wheel the cornering stiffness decreases with increasing $|\gamma|$. Note the difference in curvature left and right of the characteristics at larger values of the camber angle. To accomplish this, a split-E according to Equation 17.69 has been employed. In the complete listing of the formulae, shown later on, nondimensional parameters have been introduced. For example, the parameters in (17.70) will become $p_1 = F_{zo}p_{Ky1}$, $p_2 = F_{zo}p_{Ky2}$, and $p_3 = p_{Ky3}$, with F_{zo} denoting the nominal wheel load.

The aligning torque M_z can now be obtained by multiplying the side force F_y with the pneumatic trail t and adding the usually small (except with camber) residual torque M_{zr} (cf. Figure 17.17). We have

$$M_z = -t \cdot F_y + M_{zr} \tag{17.71}$$

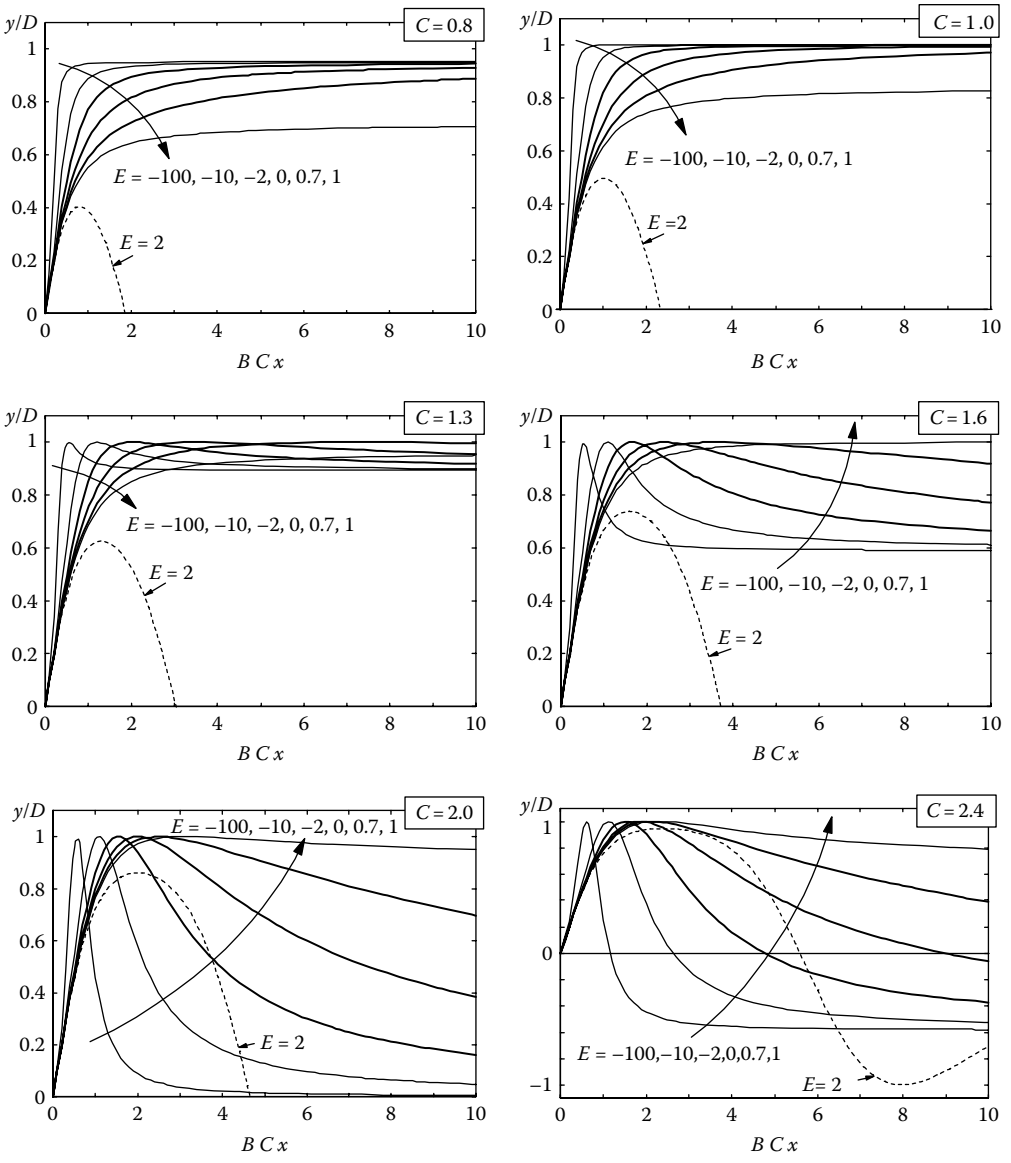


FIGURE 17.15 The influence of the shape factors C and E on the appearance of the curve according to Equation 17.64. Note that a value of $E > 1$ does not produce realistic curves.

The pneumatic trail decays with increasing side slip and is described with the aid of the cosine version of the Magic Formula as follows:

$$t(\alpha_t) = D_t \cos[C_t \arctan\{B_t\alpha_t - E_t(B_t\alpha_t - \arctan(B_t\alpha_t))\}] \tag{17.72}$$

where

$$\alpha_t = \tan\underline{\alpha} + S_{Ht} \tag{17.73}$$

The residual torque showing a similar decay

$$M_{zr}(\alpha_r) = D_r \cos\left[\arctan\left(B_r\alpha_r\right)\right] \tag{17.74}$$

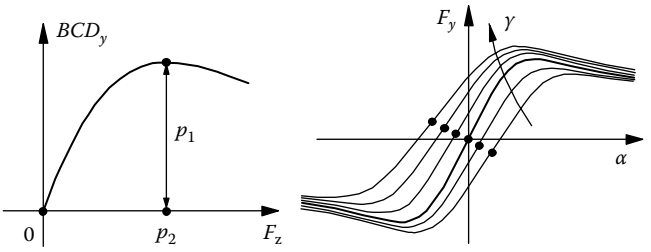


FIGURE 17.16 Cornering stiffness versus vertical load and the influence of wheel camber, Equation 17.70.

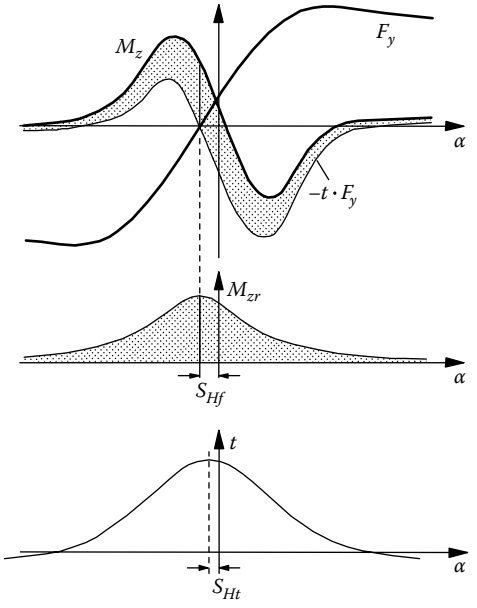


FIGURE 17.17 The aligning torque characteristic composed of a part directly attributed to the side force and a part due to the so-called residual torque (due to tire conicity and camber).

with

$$\alpha_r = \tan\alpha + S_{Hf} \tag{17.75}$$

It is seen that both parts of the moment are modeled using the Magic Formula, but instead of the sine function, the cosine function is employed. In that way, a hill-shaped curve is produced. The peaks are shifted sideways.

The residual torque is assumed to attain its maximum D_r at the slip angle where the side force becomes equal to zero. This is accomplished through the horizontal shift S_{Hf}. The peak of the pneumatic trail occurs at $\tan\alpha = -S_{Ht}$. This formulation has proven to give very good agreement with measured curves. The advantage with respect to the earlier versions, where formula (17.64) is used for the aligning torque as well, is that we have now directly assessed the function for the pneumatic trail, which is needed to handle the combined slip situation.

In Figure 17.18, the basic properties of the cosine-based curve have been indicated (subscripts of factors have been deleted again). Again, D is the peak value, C is a shape factor determining the level y_a of the horizontal asymptote, and now B influences the curvature at the peak (illustrated with

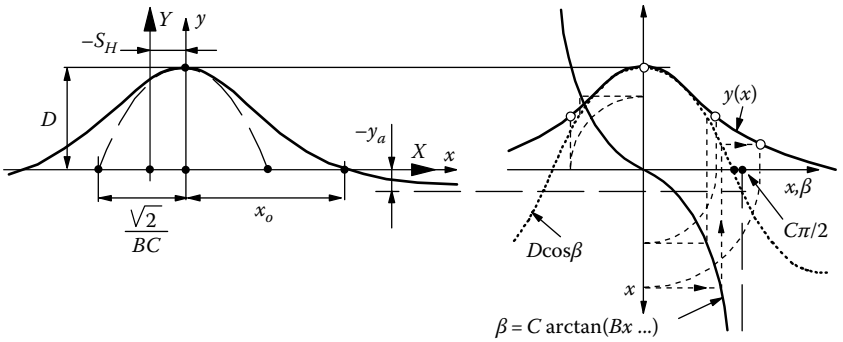


FIGURE 17.18 Curve produced by the cosine version of the Magic Formula, Equations 17.72 and 17.78). The meaning of curve parameters has been indicated.

the inserted parabola). The factor E modifies the shape at larger values of slip and governs the location x_o of the point where the curve intersects the x-axis. The following formulae hold

$$C = \frac{2}{\pi}\arccos\frac{y_a}{D} \tag{17.76}$$

$$E = \frac{Bx_o - \tan\{\pi/(2C)\}}{Bx_o - \arctan(Bx_o)} \quad (\text{if } C > 1) \tag{17.77}$$

In case of the possible presence of large camber angles (motorcycles), it may be better to use in (17.71) the side force F_y that would arise at $\gamma = 0$. Also, the side force function (17.64) and the cornering stiffness function (17.70) may be modified to better approximate large camber response for motorcycle tires (cf. De Vries [14] and [edn. 1 or 2] for a full listing of equations). We refer to [1] for the discussion of the model extension for turn slip (path curvature) also applicable in case of combined slip with braking or driving forces.

In the paper of Pacejka and Bakker [10], the tire's response to combined slip was modeled by using physically based formulae. A newer more efficient way is purely empirical. This method was developed by Michelin and published by Bayle et al. [11]. It describes the effect of combined slip on the lateral force and on the longitudinal force characteristics. Weighting functions G have been introduced that when multiplied with the original pure slip functions (17.64) produce the interaction effects of κ on F_y and of α on F_x. The weighting functions have a hill shape. They take the value one in the special case of pure slip (κ or α equal to zero). When, for example, at a given slip angle α from zero increasing brake slip is introduced, the relevant weighting function for F_y may first show a slight increase in magnitude (becoming larger than one) but will soon reach its peak after which a continuous decrease follows. The cosine version of the Magic Formula is used to represent the hill-shaped function:

$$G = D\cos[C\arctan(Bx)] \tag{17.78}$$

Here, G is the resulting weighting factor and x is either κ or $\tan\alpha$ (possibly shifted). The coefficient D represents the peak value (slightly deviating from one if a horizontal shift of the hill occurs), C determines the height of the hill's base, and B influences the sharpness of the hill. Coefficient B constitutes the main factor responsible for the shape of the weighting functions.

As an extension to the original function published by Bayle et al., the part with shape factor E will be added later on. This extension appears to improve the approximation, in particular at large levels of slip, especially in view of the strict condition that the weighting function G must remain positive for all slip conditions.

For the side force we get the following formulae:

$$F_y = G_{y\kappa} \cdot F_{yo} + S_{Vy\kappa} \tag{17.79}$$

with the weighting function now expressed such that it equals unity at $\kappa = 0$

$$G_{y\kappa} = \frac{\cos[C_{y\kappa} \arctan(B_{y\kappa}\kappa_s)]}{\cos[C_{y\kappa} \arctan(B_{y\kappa}S_{Hy\kappa})]} (>0) \tag{17.80}$$

where

$$\kappa_s = \kappa + S_{Hy\kappa} \tag{17.81}$$

and further the coefficients

$$B_{y\kappa} = r_{By1} \cos\left[\arctan\left\{r_{By2}\left(\tan\alpha - r_{By3}\right)\right\}\right] \tag{17.82}$$

$$C_{y\kappa} = r_{Cy1} \tag{17.83}$$

$$S_{Hy\kappa} = r_{Hy1} + r_{Hy2} df_z \tag{17.84}$$

$$S_{Vy\kappa} = D_{Vy\kappa} \sin[r_{Vy5} \arctan\left(r_{Vy6}\kappa\right)] \tag{17.85}$$

$$D_{Vy\kappa} = \mu_y F_z \cdot (r_{Vy1} + r_{Vy2} df_z + r_{Vy3}\gamma) \cdot \cos\left[\arctan(r_{Vy4} \tan\alpha)\right] \tag{17.86}$$

with df_z the notation for the nondimensional increment of the vertical load with respect to the (adapted) nominal load (cf. Equation 17.90 shown later on). Figure 17.19 depicts the two weighting functions displayed both as functions of α and of κ and the resulting force characteristics. In the following text, an explanation is given.

In Equation 17.79, F_{yo} denotes the side force at pure side slip obtained from Equation 17.64. The denominator of the weighting function (17.80) makes $G_{y\kappa} = 1$ at $k = 0$. The horizontal shift $S_{Hy\kappa}$ of the weighting function accomplishes the slight increase that the side force experiences at moderate braking before the peak of $G_{y\kappa}$ is reached and the decay of F_y commences. This horizontal shift may be made dependent on the vertical load. $C_{y\kappa}$ controls the level of the horizontal asymptote. If $C_{y\kappa} = 1$, the weighting function (17.80) will approach zero when $\kappa \to \pm\infty$. This would be the correct value for $C_{y\kappa}$ if κ is expected to be used in the entire range from plus to minus infinity. If this is not intended, then $C_{y\kappa}$ may be chosen different from one if $G_{y\kappa}$ is optimized with the restriction to remain positive. The factor $B_{y\kappa}$ influences the sharpness of the hill-shaped weighting function. As indicated, the hill becomes more flat (wider) at larger slip angles. Then $B_{y\kappa}$ decreases according to (17.82). When in an extreme situation α approaches 90°, that is, when $V_{cx} \to 0$, $B_{y\kappa}$ will go to zero and, consequently, $G_{y\kappa}$ will remain equal to one unless κ goes to infinity, which may easily be the case when at $V_{cx} \to 0$ the wheel speed of revolution Ω and thus the longitudinal slip velocity V_{sx} remain unequal to zero. The quantity $S_{Vy\kappa}$ is the vertical "shift," which sometimes is referred to as the κ-induced ply steer. At camber, due to the added asymmetry, the longitudinal force clearly produces a torque that creates a torsion angle comparable with a possibly already present ply-steer angle. This shift function varies with slip k as indicated in (17.49).

As illustrated in Figure 17.19, its peak value $D_{Vy\kappa}$ depends on the camber angle γ and decays with increasing magnitude of the slip angle α. Figure 17.20 presents measured data together with the fitted curves as published by Bayle et al.

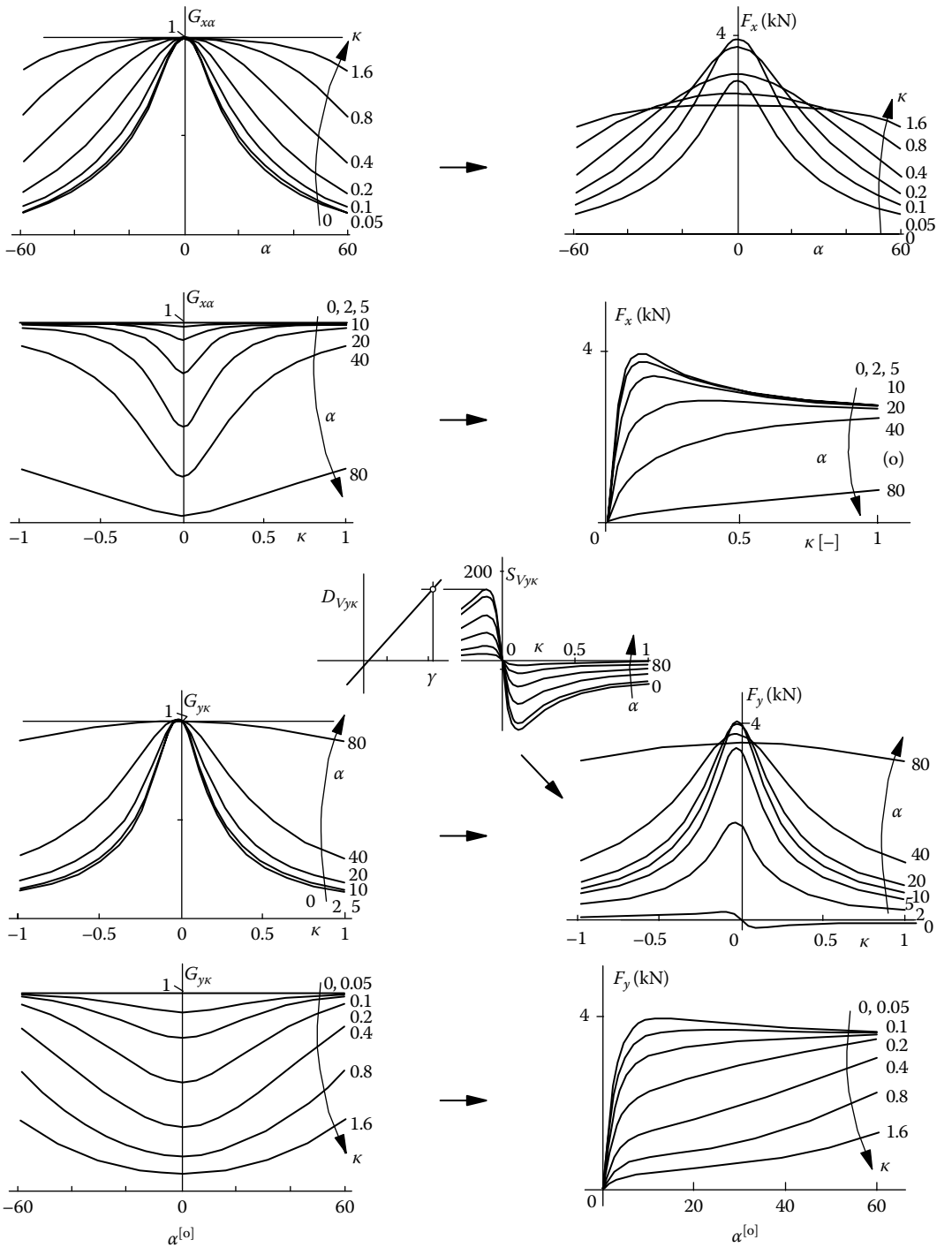


FIGURE 17.19 Nature of weighting functions and the resulting combined slip longitudinal and lateral forces, the latter also affected by the κ-induced ply-steer "vertical shift" $S_{Vy\kappa}$.

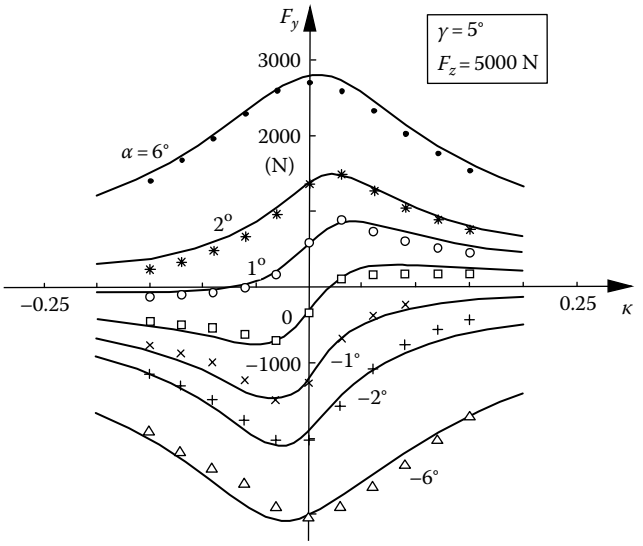


FIGURE 17.20 The combined slip side force characteristics in the presence of a camber angle. (From Bayle, P. et al., A new tyre model for vehicle dynamics simulation, *Proceedings of the Automotive Technology International*, Aachen, Germany, pp. 193–198, 1993.)

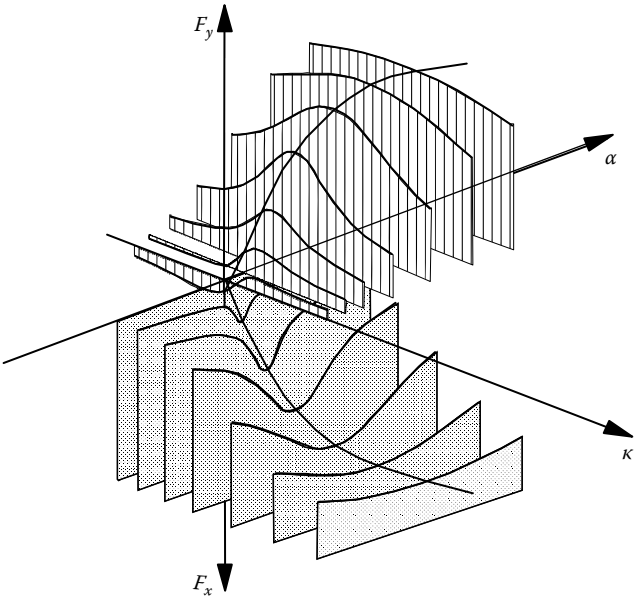


FIGURE 17.21 A 3-D graph of combined slip force characteristics.

The combined slip relations for F_x are similar to what we have seen for the side force. However, a vertical shift function was not needed to be included. In Figure 17.21, a three dimensional (3-D) graph is shown indicating the variation of F_x and F_y with both α and κ. The initial "*S*" shape of the F_y versus κ curves (at small α) due to the vertical shift function is clearly visible.

Regarding the aligning torque, physical insight is used to model the situation at combined slip. We write

$$M_z = -t(\alpha_{t,eq}) \cdot F_y + M_{zr}(\alpha_{r,eq}) + s(F_y, \gamma) \cdot F_x \tag{17.87}$$

The arguments α_t and α_r (including a shift) appearing in the functions (17.73, 17.75) for the pneumatic trail and residual torque at pure side slip are replaced by equivalent slip angles, as indicated by Equation 17.88, incorporating the effect of κ on the composite slip:

$$\alpha_{t,eq} = \sqrt{\alpha_t^2 + \left(\frac{K_{x\kappa}}{K_{y\alpha}}\right)^2 \kappa^2} \cdot \mathrm{sgn}(\alpha_t) \tag{17.88}$$

and similar for $\alpha_{r,eq}$. To approximate the same effect on the degree of sliding in the contact patch as would occur with side slip, the longitudinal slip κ is multiplied with the ratio of the longitudinal and lateral slip stiffnesses.

Besides, an extra term is introduced in (17.87) to account for the fact that a moment arm s arises for F_x as a result of camber γ and lateral tire deflection associated with F_y. This extra term may give rise to a sign change of the aligning torque in the range of braking as is observed to occur in practice.

17.4.2 Full Set of Equations

The Magic Formula model equations contain the nondimensional model parameters p, q, r, and s and, in addition, a set of scaling factors λ and spin factors ζ (camber plus turn slip, both not shown here, cf. [1]). Other parameters and variable quantities used in the equations are:

g	acceleration due to gravity
V_c	magnitude of the velocity of the wheel contact center C
$V_{cx,y}$	components of the velocity of the wheel contact center C
$V_{sx,y}$	components of slip velocity V_s (of point S)
V_r	$(= R_e\,\Omega = V_{cx} - V_{sx})$ forward speed of rolling
V_o	reference velocity $(= \sqrt{(gR_o)}$ or other specified value)
R_o	unloaded tire radius $(= r_o)$
R_e	effective rolling radius $(= r_e)$
Ω	wheel speed of revolution
ρ_z	tire normal deflection (> 0 if compression)
F_{zo}	nominal (rated) load (≥ 0)
F'_{zo}	adapted nominal load

The effect of having a tire with a different nominal load may be roughly approximated by using the scaling factor λ_{Fzo}. cf. [1], 3d ed. for influence of inflation pressure.

$$F'_{zo} = \lambda_{Fzo} F_{zo} \tag{17.89}$$

Further, we introduce the normalized change in vertical load

$$df_z = \frac{F_z - F'_{zo}}{F'_{zo}} \tag{17.90}$$

Instead of taking the slip angle α itself (in radians, from Equation 17.2) as input quantity, one may in case of very large slip angles and possibly backward running of the wheel better use the tangent of the slip angle defined as the lateral slip:

$$\alpha^* = \tan\alpha \cdot \mathrm{sgn}\, V_{cx} = -\frac{V_{cy}}{|V_{cx}|} \tag{17.91}$$

For the camber angle we introduce

$$\gamma^* = \sin\gamma \tag{17.92}$$

The longitudinal slip ratio is defined as follows:

$$\kappa = -\frac{V_{sx}}{|V_{cx}|} \tag{17.93}$$

If the forward speed V_{cx} becomes or is equal to zero, one might add a small quantity ε in the denominator of (17.91, 17.93) to avoid singularity, or, when transient slip situations occur, one should use the transient slip quantities (or deformation gradients) $\tan\alpha'$ and κ' as defined and used in the next section.

To avoid the occurrence of similar singularities in the ensuing equations due to, for example, zero velocity or zero vertical load, a small additional quantity ε (with same sign as its neighboring main quantity) will be introduced in relevant denominators like in the next equation.

For the factor cosα appearing in the equations for the aligning torque to properly handle the case of large slip angles and possibly backward running ($V_{cx} < 0$), we have defined

$$\cos'\alpha = \frac{V_{cx}}{V_c'} \tag{17.94}$$

with

$$V_c' = V_c + \varepsilon_V \tag{17.95}$$

where we may choose $\varepsilon_V = 0.1$ m/s.

For the three forces and three moments acting from road to tire and defined according to the diagram of Figure 17.1 (along and around the x-, y-, z-axes, $F_z > 0$), the equations, first those for the condition of pure slip (including camber, note that the formulation for the camber stiffness (cf. (17.114) through (17.117)) has been changed with respect to earlier publications) and subsequently those for the condition of combined slip, read: (for more details cf. [1], 3d edition).

Longitudinal force (*pure longitudinal slip*)

$$F_{xo} = D_x \sin[C_x \arctan\{B_x\kappa_x - E_x(B_x\kappa_x - \arctan(B_x\kappa_x))\}] + S_{Vx} \tag{17.96}$$

$$\kappa_x = \kappa + S_{Hx} \tag{17.97}$$

$$C_x = p_{Cx1}\ (>0) \tag{17.98}$$

$$D_x = \mu_x \cdot F_z\ (>0) \tag{17.99}$$

$$\mu_x = p_{Dx1} + p_{Dx2} df_z\ (>0) \tag{17.100}$$

$$E_x = (p_{Ex1} + p_{Ex2} df_z + p_{Ex3} df_z^2) \cdot \{1 - p_{Ex4}\,\mathrm{sgn}(\kappa_x)\} \quad (\le 1) \tag{17.101}$$

$$K_{x\kappa} = F_z \cdot (p_{Kx1} + p_{Kx2} df_z) \cdot \exp(p_{Kx3} df_z)$$

$$\left(= B_x C_x D_x = \frac{\partial F_{xo}}{\partial \kappa_x} \text{ at } \kappa_x = 0 \right) \quad \left(= C_{F\kappa} \right) \tag{17.102}$$

$$B_x = \frac{K_{x\kappa}}{C_x D_x + \varepsilon_x} \tag{17.103}$$

$$S_{Hx} = (p_{Hx1} + p_{Hx2} df_z) \tag{17.104}$$

$$S_{Vx} = F_z (p_{Vx1} + p_{Vx2} df_z) \left\{ \frac{V_{cx}}{\varepsilon_{Vx} + V_{cx}} \right\} \tag{17.105}$$

Lateral force (*pure side slip*)

$$F_{yo} = D_y \sin[C_y \arctan \{B_y \alpha_y - E_y (B_y \alpha_y - \arctan (B_y \alpha_y))\}] + S_{Vy} \tag{17.106}$$

$$\alpha_y = \alpha^* + S_{Hy} \tag{17.107}$$

$$C_y = p_{Cy1} (> 0) \tag{17.108}$$

$$D_y = \mu_y \cdot F_z \tag{17.109}$$

$$\mu_y = \frac{p_{Dy1} + p_{Dy2} df_z}{1 + p_{Dy3} \gamma^{*2}} (> 0) \tag{17.110}$$

$$E_y = (p_{Ey1} + p_{Ey2} df_z)\{1 + p_{Ey5} \gamma^{*2} - (p_{Ey3} + p_{Ey4} \gamma^*) \operatorname{sgn}(\alpha_y)\} \ (\leq 1) \tag{17.111}$$

$$\begin{gathered} K_{ya} = p_{Ky1} F'_{zo} \sin \frac{p_{Ky4} \arctan\{F_z / ((p_{Ky2} + p_{Ky5} \gamma^{*2}) F'_{zo})\}}{1 + p_{Ky3} \gamma^{*2}} \\ (= B_y C_y D_y = \frac{\partial F_{yo}}{\partial \alpha_y} \text{ at } \alpha_y = 0) \text{ (if } \gamma = 0 := K_{y\alpha o} = C_{F\alpha}) \text{ (usually : } p_{Ky4} = 2) \end{gathered} \tag{17.112}$$

$$B_y = \frac{K_{ya}}{C_y D_y + \varepsilon_y} \tag{17.113}$$

$$S_{Hy} = p_{Hy1} + p_{Hy2} df_z + \frac{K_{y\gamma o} \gamma^* - S_{Vy\gamma}}{K_{y\alpha} + \varepsilon_K} \tag{17.114}$$

$$S_{Vy\gamma} = F_z \cdot (p_{Vy3} + p_{Vy4} df_z) \gamma^* \tag{17.115}$$

$$S_{Vy} = F_z \cdot (p_{Vyl} + p_{Vy2} df_z) + S_{Vy\gamma} \tag{17.116}$$

$$K_{y\gamma o} = F_z (p_{Ky6} + p_{Ky7} df_z) \left(=\sim \frac{\partial F_{yo}}{\partial \gamma} \text{ at } \alpha = \gamma = 0 \right) (= C_{F\gamma}) \tag{17.117}$$

Aligning torque (*pure side slip*)

$$M_{zo} = M'_{zo} + M_{zro} \tag{17.118}$$

$$M'_{zo} = -t_o \cdot F_{yo} \tag{17.119}$$

$$t_o = t(\alpha_t) \ = D_t \cos[C_t \arctan\{B_t\alpha_t - E_t(B_t\alpha_t - \arctan(B_t\alpha_t))\}] \cdot \cos'\alpha \tag{17.120}$$

$$\alpha_t = \alpha^* + S_{Ht} \tag{17.121}$$

$$S_{Ht} = q_{Hz1} + q_{Hz2}df_z + (q_{Hz3} + q_{Hz4}df_z)\gamma^* \tag{17.122}$$

$$M_{zro} = M_{zr}(\alpha_r) = D_r \cos[C_r \arctan(B_r\alpha_r)] \cdot \cos'\alpha \tag{17.123}$$

$$\alpha_r = \alpha^* + S_{Hf}(= \alpha_f) \tag{17.124}$$

$$S_{Hf} = \frac{S_{Hy} + S_{Vy}}{K'_{y\alpha}} \tag{17.125}$$

$$K'_{y\alpha} = K_{ya} + \varepsilon_K \tag{17.126}$$

$$B_t = \left(q_{Bzl} + q_{Bz2}df_z + q_{Bz3}df_z^2\right) \cdot \left(1 + q_{Bz5}\left|\gamma^*\right| + q_{Bz6}\gamma^{*2}\right) \quad (>0) \tag{17.127}$$

$$C_t = q_{Cz1} (>0) \tag{17.128}$$

$$D_{to} = F_z \cdot \left(\frac{R_o}{F'_{zo}}\right) \cdot \left(q_{Dz1} + q_{Dz2}df_z\right) \cdot \operatorname{sgn} V_{cx} \tag{17.129}$$

$$D_t = D_{to} \cdot (1 \ + q_{Dz3}\left|\gamma^*\right| + q_{Dz4}\gamma^{*2}) \tag{17.130}$$

$$E_t = (q_{Ez1} + q_{Ez2}df_z + q_{Ez3}df_z^2)\left\{1 + (q_{Ez4} + q_{Ez5}\gamma^*)\frac{2}{\pi}\operatorname{archan}(B_tC_t\alpha_t)\right\} \quad (\leq 1) \tag{17.131}$$

$$B_r = q_{Bz9} + q_{Bz10}B_yC_y \ (\text{preferred} : q_{Bz9} = \ 0) \tag{17.132}$$

$$C_r = 1 \tag{17.133}$$

$$D_r = F_zR_o\left\{(q_{Dz6} + q_{Dz7}df_z) + (q_{Dz8} + q_{Dz9}df_z)\gamma^* + (q_{Dz10} + q_{Dz11}df_z)\gamma^*\left|\gamma^*\right|\right\}\cos'\alpha \cdot \operatorname{sgn} V_{cx} \tag{17.134}$$

$$K_{zao} = D_{to}K_{y\alpha,\gamma=0} \quad \left(=\sim \frac{-\partial M_{zo}}{\partial \alpha_y} \text{ at } \alpha_y = \gamma = 0\right) \quad (= C_{Ma}) \tag{17.135}$$

$$K_{z\gamma o} = F_zR_o(q_{Dz8} + q_{Dz9}df_z) - D_{to}K_{y\gamma o}\left(=\sim \frac{\partial M_{zo}}{\partial \gamma} \text{ at } \alpha = \gamma = 0\right) \quad (= C_{M\gamma}) \tag{17.136}$$

Longitudinal force (*combined slip*)

$$F_x = G_{x\alpha} \cdot F_{xo} \tag{17.137}$$

$$G_{x\alpha} = \cos \frac{C_{x\alpha} \arctan\left\{B_{x\alpha}\alpha_s - E_{x\alpha}(B_{x\alpha}\alpha_s - \arctan(B_{x\alpha}\alpha_s))\right\}}{G_{x\alpha o}(>0)} \tag{17.138}$$

$$G_{x\alpha o} = \cos[C_{x\alpha}\arctan\{B_{x\alpha}S_{Hx\alpha} - E_{x\alpha}(B_{x\alpha}S_{Hx\alpha} - \arctan(B_{x\alpha}S_{Hx\alpha}))\}] \tag{17.139}$$

$$\alpha_s = \alpha^* + S_{Hx\alpha} \tag{17.140}$$

$$B_{x\alpha} = \left(r_{Bx1} + r_{Bx3}\gamma^{*2}\right)\cos[\arctan(r_{Bx2}\kappa)\ (>0) \tag{17.141}$$

$$C_{x\alpha} = r_{Cx1} \tag{17.142}$$

$$E_{x\alpha} = r_{Ex1} + r_{Ex2}df_z (\le 1) \tag{17.143}$$

$$S_{Hx\alpha} = r_{Hx1} \tag{17.144}$$

Lateral force (*combined slip*)

$$F_y = G_{y\kappa} \cdot F_{yo} + S_{Vyx} \tag{17.145}$$

$$G_{y\kappa} = \cos \frac{C_{y\kappa}\arctan\left\{B_{y\kappa}\kappa_S - E_{y\kappa}\kappa_S - \arctan(B_{y\kappa}\kappa_S)\right\}}{G_{y\kappa o}(>0)} \tag{17.146}$$

$$G_{y\kappa o} = \cos[C_{y\kappa}\arctan\{B_{y\kappa}S_{Hy\kappa} - E_{Hy\kappa}(B_{y\kappa}S_{Hy\kappa} - \arctan(B_{y\kappa}S_{Hy\kappa}))\} \tag{17.147}$$

$$k_s = k + S_{Hy\kappa} \tag{17.148}$$

$$B_{y\kappa} = \left(r_{By1} + r_{By4}\gamma^{*2}\right)\cos[\arctan\{r_{By2}(\alpha^* - r_{By3})\}]\ (>0) \tag{17.149}$$

$$C_{y\kappa} = r_{Cy1} \tag{17.150}$$

$$E_{y\kappa} = r_{Ey1} + r_{Ey2}df_z (\le 1) \tag{17.151}$$

$$S_{Hy\kappa} = r_{Hy1} + r_{Hy2}df_z \tag{17.152}$$

$$S_{Vy\kappa} = D_{Vy\kappa}\sin[r_{Vy5}\arctan(r_{Vy6}\kappa)] \tag{17.153}$$

$$D_{Vy\kappa} = \mu_y F_z \cdot (r_{Vy1} + r_{Vy2}df_z + r_{Vy3}\gamma^*) \cdot \cos[\arctan(r_{Vy4}\alpha^*) \tag{17.154}$$

Normal load

$$F_z = P_{z1} \cdot (F'_{zo}) \cdot \rho_z \quad (\geq 0)$$

$$\left(C_{Fz} = \frac{\partial F_z}{\partial \rho_z} = \frac{P_{z1} F'_{zo}}{R_0} \right) \tag{17.155}$$

Overturning couple

$$M_x = F_z R_o \cdot \left(\frac{q_{sx1} - q_{sx2}\gamma^* + q_{sx3} F_y}{F'_{zo}} \right) \tag{17.156}$$

Rolling resistance moment

$$M_y = -F_z R_o \cdot q_{sy1} \arctan\left(\frac{V_r}{V_o} \right) + \frac{q_{sy2} F_x}{F'_{zo}} \tag{17.157}$$

Aligning torque (*combined slip*)

$$M_z = M'_z + M_{zr} + s \cdot F_x \tag{17.158}$$

$$M_z = -t \cdot F'_y \tag{17.159}$$

$$t = t(\alpha_{t,eq}) = D_t \cos[C_t \arctan\{B_t \alpha_{t,eq} - E_t (B_t \alpha_{t,eq} - \arctan(B_t \alpha_{t,eq}))\}] \cdot \cos' \alpha \tag{17.160}$$

$$F'_y = F_y - S_{Vyk} \tag{17.161}$$

$$M_{zr} = M_{zr}(\alpha_{r,eq}) = D_r \cos[C_r \arctan(B_r \alpha_{r,eq})] \tag{17.162}$$

$$s = R_0 \cdot \left\{ s_{sz1} + s_{sz2} \left(\frac{F_y}{F'_{z0}} \right) + (s_{sz3} + s_{sz4} df_z) \gamma^* \right\} \tag{17.163}$$

$$\alpha_{t,eq} = \sqrt{\alpha_t^2 + \left(\frac{K_{xk}}{K'_{ya}} \right)^2 k^2} \cdot \operatorname{sgn}(\alpha_t) \tag{17.164}$$

$$\alpha_{r,eq} = \sqrt{\alpha_r^2 + \left(\frac{K_{xk}}{K'_{ya}} \right)^2 k^2} \cdot \operatorname{sgn}(\alpha_r) \tag{17.165}$$

In the textbook [1], a set of parameter values have been listed for a specific tire. Examples of computed characteristics compared with experimentally assessed curves for both pure slip and combined slip conditions have been presented. The formulae appear to produce very good agreement. The effect of having turn slip is modeled and some calculated characteristics have been presented for a set of hypothetical model parameters. Also, the possibility to define the effects of conicity and ply steer as responses to equivalent camber and equivalent slip angles has been discussed. A possibly better description than (17.156) for the overturning couple of truck, car, and motorcycle tires due to side slip and camber has been introduced.

17.5 FIRST-ORDER TIRE LAG

For relatively low-frequency and large-wavelength transient and oscillatory vehicle motions, tire inertia and the effect of the finite length of the contact patch may be neglected or taken care of in an approximative manner. In [1], a thorough treatment has been given of the out-of-plane stretched string tire model (based on the early work of von Schlippe [15]) together with a number of extended and approximated models. In one of these approximate models, the effect of the finite length of the contact patch has been ignored. For the aligning torque, the model includes the effects of tread width and gyroscopic couple.

The present section deals with the further development of this type of single-contact-point model that in its simplest form is very popular in vehicle dynamics studies. Both in-plane and out-of-plane models will be discussed for small slip, linear and for large slip, nonlinear conditions. The concept of the relaxation length is central in the model structure. The development of the single-contact-point models follows an essentially different and much simpler line compared with the theoretical approach on which the string model is based. Because of its simplicity, it is possible to enhance the model to cover the full nonlinear combined slip range including rolling from standstill or even change direction from forward to backward rolling. Camber and turn slip may be included.

With the linear and nonlinear models, various vehicle dynamics problems may be approximately analyzed such as the shimmy phenomenon, transient vehicle motions with oscillatory steer inputs, motion over undulated road surfaces at side slip and camber and at controlled braking, steering vibrations induced by wheel imbalance, and tire out-of-roundness. In [1] a number of applications addressing these problems have been studied.

17.5.1 Model Development

The model consists of a contact patch (point) that is suspended with respect to the wheel rim by a longitudinal (circumferential) and a lateral spring. These springs represent the compliance of the tire carcass. The contact point may move (slip) with respect to the ground in lateral and longitudinal directions. Through this relative motion the side force, longitudinal force, and aligning torque are generated. To determine these forces and moments, contact patch lateral slip (possibly including ply steer) and longitudinal slip are defined. In addition, contact line curvature, due to camber possibly including conicity and due to turn slip, is thought to occur. These contact patch slip quantities may then be used as input in the steady-state tire slip model, for example, the Magic Formula, to calculate the transient force and moment variation that act upon the contact patch.

17.5.2 Linear Model

Figure 17.22 depicts the model in top view. At the instant considered, the wheel slip point S (attached to the wheel rim at a level near the road surface) and the contact point S' are defined to

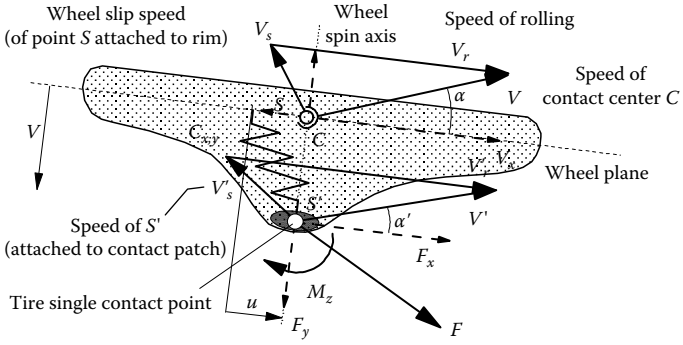


FIGURE 17.22 Single-contact-point tire model showing lateral and longitudinal carcass deflections, u and v (top view).

be located in the plane through the wheel axis and normal to the road. These points (which may be thought to lie on two parallel slip circles) move over the road surface with the wheel and contact patch slip velocities, respectively. In the figure, the x and y components of the slip velocities have been shown as negative quantities. The difference of the velocities of the two points causes the carcass springs to deflect. Consequently, the time rates of change of the longitudinal and lateral deflections u and v read

$$\frac{du}{dt} = -(V_{sx} - V'_{sx}) \tag{17.166}$$

and

$$\frac{dv}{dt} = -(V_{sy} - V'_{sy}) \tag{17.167}$$

If we assume small values of slip, we may write for the side force acting from road to contact patch with $C_{F\alpha}$ denoting the cornering or side slip stiffness:

$$F_y = C_{F\alpha}\alpha' = -C_{F\alpha}\frac{V'_{sy}}{|V_x|} \tag{17.168}$$

It is assumed here that the difference between the wheel center longitudinal velocity V_x and the longitudinal velocity V_{cx} of the contact center is negligible:

$$V_{cx} \approx V_x \tag{17.169}$$

Consequently, we may employ here V_x in the denominator. With the lateral tire stiffness at road level C_{Fy}, we have for the elastic internal force that balances the side slip force

$$F_y = C_{Fy}v \tag{17.170}$$

and we can write for Equation 17.167 with Equations 17.168 and 17.170 after having introduced the relaxation length for side slip σ_a:

$$\sigma_\alpha = \frac{C_{F\alpha}}{C_{Fy}} \tag{17.171}$$

The differential equation for the lateral deflection due to side slip v_α (later on we will also have a lateral deflection due to camber γ) is

$$\frac{dv_\alpha}{dt} + \frac{1}{\sigma_\alpha}|V_x|v_\alpha = |V_x|\alpha = -V_{sy} \tag{17.172}$$

where α is the wheel slip angle: $\alpha \approx -V_{sy}/|V_x|$. The side force is obtained by multiplying v_α with C_{Fy}.

In a similar way we can deal with the longitudinal force response. With the longitudinal tire stiffness C_{Fx} at road level and the longitudinal slip stiffness C_{Fk}, we obtain for the relaxation length σ_k

$$\sigma_k = \frac{C_{Fk}}{C_{Fx}} \tag{17.173}$$

and we can derive with Equation 17.166 the differential equation for the fore and aft deflection u:

$$\frac{du}{dt} + \frac{1}{\sigma_k}|V_x|u = |V_x|K = -V_{sx} \tag{17.174}$$

with k the longitudinal wheel slip ratio: $k \approx -V_{sx}/|V_x|$. The longitudinal force may be obtained by multiplying u with the stiffness C_{Fx}.

Next, we consider the wheel camber as input. For a suddenly applied camber angle γ (about the line of intersection at ground level), we assume that a contact line curvature and thus the camber thrust $C_{F\gamma}\gamma$ are immediately felt at the contact patch. As a reaction, a contact patch side slip angle α' is developed that builds up the lateral carcass deflection $v = v_\gamma$. Again Equation 17.170 applies. The side force that acts on the wheel now becomes

$$F_y = C_{FY}V_\gamma = C_{F\gamma}\gamma + C_F\alpha' \tag{17.175}$$

With wheel side slip kept equal to zero, $V_{sy} = 0$, and $V'_{sy} = -|V_x|\alpha'$, Equation 17.167 can be written in the form:

$$\frac{dv_\gamma}{dt} + \frac{1}{\sigma_\alpha}|V_x|v_\gamma = \frac{C_{F\gamma}}{C_{F\alpha}}|V_x|\gamma \tag{17.176}$$

This equation shows that according to this simple model, the camber force relaxation length σ_γ is equal to the relaxation length for side slip σ_α. This theoretical result is substantiated by careful camber step response experiments, performed by Higuchi [16] on a flat plank test rig. A similar equation results for the total spin φ including turn slip and camber, [1].

The forces and moment may be obtained from the deflections u and v by first assessing the transient slip quantities α', k', and γ' and from these with the slip stiffnesses the forces and moment.

According to the adopted steady-state model, Equation 17.158, the moment response to camber is the sum of the residual torque, M_{zr}, assumedly mainly due to finite tread width, and $-t_\alpha F_y$, assumedly caused by camber-induced side slip. A first-order approximation for the response of M_{zr} with the short relaxation length equal to half the contact length a may be employed. Here, however, we suffice with the assumption that the moment due to tread width responds instantaneously to camber.

For the linear, small slip condition we find

$$\alpha' \approx \tan\alpha' = \frac{v_\alpha}{\sigma_\alpha}, \quad F_{y\alpha} = C_{F\alpha}\alpha' \quad M_{Z\alpha} = -C_{M\alpha}\alpha' = -t_\alpha F_{y\alpha} \tag{17.177}$$

$$k' = \frac{u}{\sigma_k}, \quad F_x = C_{Fk}k' \tag{17.178}$$

$$\gamma' = \frac{C_{F\alpha}}{C_{F\gamma}}\frac{v_\gamma}{\sigma_\alpha}, \quad F_{y\gamma} = C_{F\gamma}\gamma', \quad M_{zry} = (C_{M\gamma} + t_\alpha C_{F\gamma})\gamma \tag{17.179}$$

The pneumatic trail due to side slip is denoted here with t_α. The total aligning torque becomes

$$M_z = -t_\alpha(F_{y\alpha} + F_{y\gamma}) + M_{zr\gamma} \tag{17.180}$$

In an alternative model, used for motorcycle dynamic studies, the terms with $t_\alpha C_{Fy}$ or $t_\alpha F_{y\gamma}$ in (17.179, 17.180) are omitted (cf. discussion in [1] following Equation 7.40).

The Equations (17.172, 17.174, 17.176) may be written in terms of the transient slip quantities. For example, we may express v_α in terms of α' by using the first equation of (17.177). Insertion in (17.172) gives

$$\sigma_\alpha \frac{d\alpha'}{dt} + |V_x|\alpha' = |V_x|\alpha = -V_{sy} \tag{17.181}$$

If we recognize the fact that the relaxation length is a function of the vertical load and if the average slip angle is unequal to zero, an additional term shows up in the linearized equation (variations of α' and of $F_z(t)$ are small) that results from the differentiation of $v_\alpha = \sigma_\alpha \alpha'$ with respect to time. Then, Equation 17.172 becomes

$$\alpha' \frac{d\alpha'}{dF_z} + \left(|V_x| + \frac{d\alpha'}{dF_z}\frac{dF_z}{dt} \right) \alpha' = |V_x|\alpha = -V_{sy} \tag{17.182}$$

Obviously, when using Equation 17.181, a response to a variation of the vertical load cannot be expected. If the load varies, Equation 17.181 is inadequate and the original Equation 17.172 or the corresponding Equation 17.182 should be used.

With (17.170) Equation 17.172 may be written directly in terms of F_y. If we may consider the carcass lateral stiffness C_{Fy} virtually independent of the wheel load F_z, we obtain by using Equation 17.171:

$$\sigma_\alpha \frac{dF_y}{dt} + |V_x| F_y = |V_x| F_{yss} \tag{17.183}$$

Since we have the same relaxation length for both the responses to side slip and camber, this equation appears to hold for the combined linear response to the inputs α and γ or φ. In the right-hand member, F_{yss} denotes the steady-state (instantaneous) response to these inputs and possibly a changing vertical load at a given slip condition. Multiplication with the pneumatic trail produces the moment $-M'_z$ (if $\gamma = \varphi = 0$). A similar differential equation may be written for the F_x response to k.

Equations 17.172 and 17.174 have been written in the form (V_x not in denominator and $V_{sx,y}$ used as right-hand member) that makes them suitable for simulations of stopping and starting from zero speed occurrences. At speed $V_x = 0$, Equation 17.174 turns into an integrator: $u = -\int V_{sx} dt$. With Equation 17.178 the longitudinal force becomes $F_x = C_{Fk} k' = C_{Fk} u/\sigma_k$, which with (17.175) is equal to $C_{Fx} u$. This is the correct expression for the tire that at standstill acts like a longitudinal or tangential spring. When the wheel starts rolling, the tire gradually changes into a damper with rate $C_{Fk}/|V_x|$. Figure 17.23 depicts a corresponding mechanical model with spring and damper in series. It shows that at low speed the damper becomes very stiff and the spring dominates. At higher forward

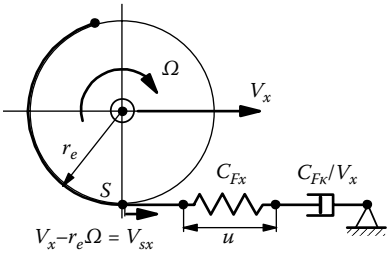


FIGURE 17.23 Mechanical model of transient tangential tire behavior.

velocities the spring becomes relatively stiff and the damper part dominates the behavior of the tire. A similar model may be drawn for the transient lateral behavior. It may be noted that Equation 17.183 is not suited for moving near or at $V_x = 0$. In [1] the use of the transient models for the response to lateral and longitudinal wheel slip speed at and near zero speed is demonstrated.

17.5.3 Semi-Nonlinear Model

For the extension of the linear theory to cover the nonlinear range of the slip characteristics, it may be tempting to employ Equation 17.183 and use the instantaneous nonlinear force response as input in the differential equation. The input steady-state side force is calculated, for example, with the Magic Formula, using the current wheel slip angle α. This method, however, may lead to incorrect results as due to the phase lag in side force response; the current (varying) wheel load may not correspond to the calculated magnitude attained by the side force. In limit conditions, the tire may then be predicted to be still in adhesion, while in reality full sliding occurs. A better approach is the utilization of the original Equation 17.172 and the calculation of the side force afterward by using the resulting transient slip angle α' (17.177) as input in the Magic Formula.

In general, we have the three Equations 17.172, 17.174, 17.176 and the first equations of (17.177) through (17.179) producing α', κ', and γ' or φ' that are used as input in the nonlinear force and moment functions (γ or φ directly in the expressions for M_r), for example, the equations of the Magic Formula tire model (Section 17.4):

$$F_x = F_x(k', \alpha', F_z) \tag{17.184}$$

$$F_y = F_y(\alpha', \gamma', K', F_z) \tag{17.185}$$

$$M_z' = -t_\alpha F_y \tag{17.186}$$

$$M_{zr} = M_{zr}(\gamma, \alpha', K', F_z) \tag{17.187}$$

$$M_z = M_z' + M_{zr} + s \cdot F_x \tag{17.188}$$

where if required, γ may be replaced by φ as the spin argument, cf. [1], edn.3.

This nonlinear model is straightforward and is often used in transient or low-frequency vehicle motion simulation applications. Starting from zero speed or stopping to standstill is possible. However, as has been mentioned before, at V_x equal or close to zero, Equations 17.172 and 17.174 act as integrators of the slip speed components $V_{sx,y}$, which may give rise to possibly very large deflections. The limitation of these deflections may be accomplished by making the derivatives of the deflections u and v equal to zero, if (1) the forward wheel velocity has become very small ($<V_{low}$) and (2) the deflections take values larger than physically possible (cf. [1]).

Another extreme situation is the condition at wheel lock. At steady state, Equation 17.174 reduces to

$$\frac{1}{\sigma_k}|V_x|u = -V_x \tag{17.189}$$

which indicates that the deflection u according to the semi-linear theory becomes as large as the relaxation length σ_k. Avoiding the deflections from becoming too large, which is of importance at, for example, repetitive braking, calls for an enhanced nonlinear model. Another shortcoming of the model that is to be tackled concerns the experimentally observed property of the tire that its relaxation length depends on the level of slip. At higher levels of side slip, the tire shows a quicker response to additional changes in side slip. This indicates that the relaxation lengths decrease with increasing slip.

17.5.4 Fully Nonlinear Model

With the first equation of (17.177), we may write the original Equation 17.172 as follows:

$$\frac{dv_\alpha}{dt} + |V_x| \tan\alpha' = |V_x| \tan\alpha = -V_{sy} \tag{17.190}$$

Writing this equation entirely in terms of the transient slip angle by using (17.170) with $v = v_\alpha$ and remembering that $F_y = F_y(\alpha', F_z)$ directly yields

$$\frac{1}{C_{Fy}} \frac{\partial F_y}{\partial \tan\alpha'} \frac{d\tan\alpha'}{dt} + |V_x| \tan\alpha' = -V_{sy} - \frac{1}{C_{Fy}} \frac{\partial F_y}{\partial F_z} \frac{dFz}{dt} \tag{17.191}$$

The additional input dF_z/dt requires information of the slope $\partial F_y/\partial F_z$ at given values of the slip angle. If the vertical load remains constant, the last term vanishes and we have the often used equation of the *restricted fully nonlinear model*:

$$\sigma_\alpha \frac{d\tan\alpha'}{dt} + |V_x| \tan\alpha' = -V_{sy} \tag{17.192}$$

with

$$\sigma_\alpha = \frac{1}{C_{Fy}} \frac{\partial F_y}{\partial \tan\alpha'} \tag{17.193}$$

If we consider an average slip angle α_o and a small variation $\tilde{\alpha}$ of tanα and the corresponding lateral slip velocities, Equation 17.192 becomes after having subtracted the average part:

$$\sigma_\alpha \frac{d\tilde{\alpha}'}{dt} + |V_x| \tilde{\alpha}' = -\tilde{V}_{sy} \tag{17.194}$$

which indicates that the structure of (17.192) is retained and that σ_α (17.193) represents the actual relaxation length of the linearized system at a given load and slip angle. Obviously, the relaxation length is associated with the slope of the side force characteristic. It also shows that the relaxation length becomes negative beyond the peak of the side force characteristic that makes the solution of (17.194) but also of the original Equation 17.190 unstable if the point of operation lies in that range of side slip. We may, however, limit σ_α downward to avoid both instability and excessive computation time: $\sigma_\alpha = \max(\sigma_\alpha, \sigma_{min})$. The transient response of the variation of the force proceeds in proportion with the variation of $\tilde{\alpha}'$ as $\tilde{F}_y = (\partial F_y/\partial \tan\alpha)\tilde{\alpha}'$.

When Equation 17.192 is used, an algebraic loop does not occur as would be the case when a general equation, the nonlinear version of Equation 17.172, for the lateral deflection v is employed (cf. [1]). The relaxation length σ_α can be directly determined from the already available tanα'. However, since the last term of (17.191) has been omitted, Equation 17.192 has become insensitive to F_z variations.

A similar function and differential equation can be derived for the transient response to longitudinal slip. We obtain for the relaxation length

$$\sigma_k = \frac{1}{C_{Fx}} \frac{\partial F_x}{\partial k'} \tag{17.195}$$

It is of interest to note that the extreme case of wheel lock can now be handled correctly with $u = F_x/C_{Fx}$ (cf. [1]).

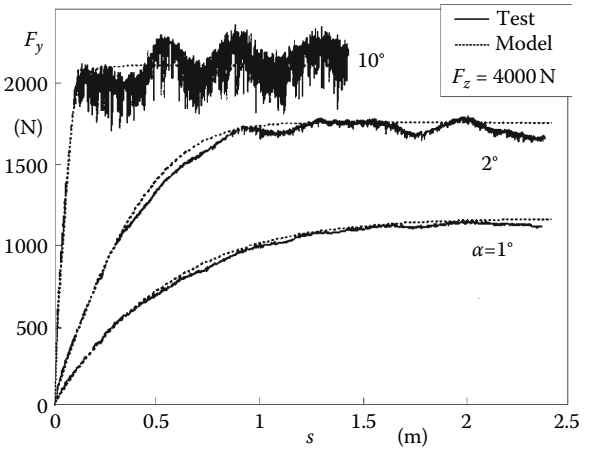


FIGURE 17.24 Side force response to small and large step change in slip angle as assessed by flat plank experiments and computed with the model defined by Equation 17.190.

Figures 17.24 through 17.26 present results obtained by Higuchi [16] for a passenger car 205/60R15 tire tested on a flat plank machine (TU Delft). Figure 17.24 shows the response of the side force to step changes in slip angle at the nominal load of 4000 N. The diagram clearly indicates the difference in behavior at the different levels of side slip with a very rapid response occurring at $\alpha = 10°$. The fully nonlinear model according to Equation 17.190, in conjunction with Equation 17.185, gives satisfactory agreement. The measured response curves have been corrected for the side force variation that for the not fully symmetric and regular tire arises already at zero slip angle. The test is performed by loading the tire after the slip angle (steer angle) has been applied. Subsequently, the plank is moved. The aligning torque behaves in a similar manner (not shown) although a little initial delay in response occurs as correctly predicted by the string model (cf. [1]). At larger slip angles (beyond the level where the M_z peak occurs), the aligning torque first shows a peak in the response after which the moment decays to its steady-state value. This behavior is nicely followed by the model [1]. Figure 17.25 shows the side force response to small increments in slip angle. The experiment is conducted by, after having stopped the plank motion, steering the wheel half a degree further and continuing the forward motion. It was first ascertained that the response of the side force to a step steer input was hardly influenced by having first lifted the tire from the road surface or not. Of course, the aligning moment response, in the latter case, is quite different because of initial torsion about the vertical axis. The resulting side

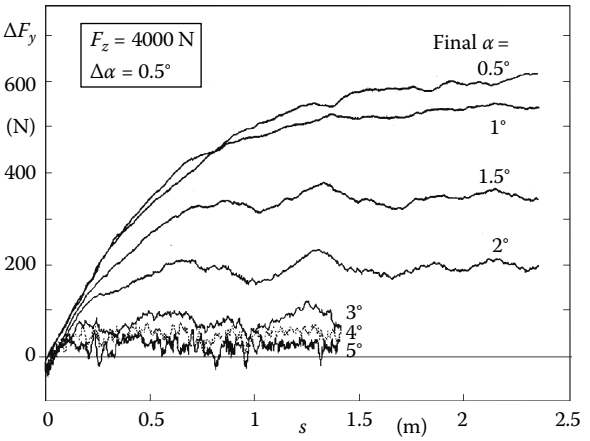


FIGURE 17.25 Side force response to small increment of slip angle $\Delta\alpha$ at different levels of slip angle indicating the decrease of relaxation length at increasing side slip level that is in agreement with Equation 17.193.

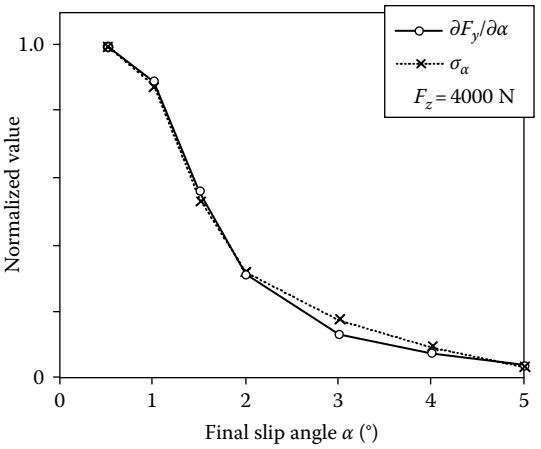


FIGURE 17.26 Correspondence between F_y versus α slope and relaxation length σ_α as assessed by experiments.

force responses as depicted in Figure 17.25 have been compared with the exponential response of the side force increment as predicted by the solution of Equation 17.194 with (17.193). Fitting with a least square procedure gave the relaxation length σ_α as presented in Figure 17.26. The resulting variation of σ_α with slip angle level is compared with the slope of the steady-state side force characteristic, both normalized making their values equal to unity at vanishing slip angle.

The excellent agreement supports the theoretical findings. Higuchi also investigated the responses to changes in camber angle and to a limited extent also to the related turn slip. The force response to a change in camber exhibits a peculiar feature. It turns out that directly after the wheel is cambered (about the line of intersection), a side force is developed instantaneously. This, obviously, is caused by the nonsymmetric distortion of the cross section of the lower part of the tire. This initial "non-lagging" part of the side force has been discussed elaborately in Ref. [1].

17.5.5 Gyroscopic Couple

In [1] the gyroscopic couple about the vertical axis that arises as a result of the time rate of change of the average tire tilt deflection angle has been introduced. This angle is considered to be proportional with the lateral tire deflection v or the side force F_y. The moment may be written in terms of the deflection rate:

$$M_{z,gyr} = c_{gyr} m_t r_e \Omega \frac{dv}{dt} \tag{17.196}$$

where at free rolling the wheel speed of revolution equals the forward velocity divided by the effective rolling radius of the tire. In general, we have

$$\Omega = \frac{V_x - V_{sx}}{r_e} \tag{17.197}$$

For a radial ply steel-belted car tire, the nondimensional coefficient c_{gyr} has been estimated to take the value 0.5. The quantity m_t represents the mass of the tire. Extending the expression (17.188) or Equation 17.158 yields for the total aligning moment:

$$M_z = M'_z + M_{zr} + sF_x + M_{z,gyr} \tag{17.198}$$

The gyroscopic couple has a large damping effect on steering vibrations of front wheels, notably on unstable shimmy oscillations (cf. [1]).

17.5.6 Enhanced Nonlinear Transient Tire Model

A totally different approach to model the transient rolling properties of the tire is based on the separation of contact patch slip properties and carcass compliance not through the use of relaxation lengths but by incorporating the carcass springs in the model explicitly. The contact patch may be given some inertia to facilitate the computational process (computational causality). This has the drawback that a relatively high natural frequency is introduced, possibly making the computation slower. We may, however, employ alternative methods to avoid the inclusion of the small mass. The model to be discussed automatically accounts for the property that the lag in the response to wheel slip and load changes diminishes at higher levels of slip that in the previous section was realized by decreasing the relaxation length. This latter approach, however, appeared to possibly suffer from computational difficulties if the model has to also account for the transient response to wheel load variations. Also, the response to combined slip was less easy to model. In developing the enhanced model, we should, however, try to maintain the nice feature of the relaxation length model to adequately handle the simulation at speeds near or equal to zero.

Figure 17.27 depicts the structure of the enhanced transient model. The contact patch can deflect in circumferential and lateral direction with respect to the lower part of the wheel rim. Only translations are allowed to ensure that the slip angle seen by the contact patch at steady state is equal to that of the wheel plane. As mentioned before, a mass point is thought to be attached to the contact patch. That mass point coincides with point S^*, the velocity of which constitutes the slip speed of the contact point. This slip velocity is used, together with the (assumedly at the contact patch detected) path curvature due to turn slip and wheel camber, to compute the forces F_x, F_y and the moment M_z that act from road to contact patch. We may add a simple relaxation length model to enable computations near zero speed. The model was first employed by Van der Jagt et al. [17] and later generalized by Pacejka and Besselink [18].

With mass m_c and carcass stiffnesses $c_{cx,y}$ and damping ratios $k_{cx,y}$ introduced, the equations of motion for the contact patch mass point with longitudinal and lateral speed components $V^*_{sx,y}$ read (yaw rate term in accelerations disregarded):

$$m_c\dot{V}^*_{sx} + k_{cx}\dot{u} + c_{cx}u = F_x(\kappa',\alpha',F_z) \tag{17.199}$$

$$m_c\dot{V}^*_{sy} + k_{cy}\dot{v} + c_{cy}v = F_y(\alpha',\kappa',\gamma,F_z) - F_{y,NL} \tag{17.200}$$

The forces acting from ground to contact patch shown at the right-hand sides are computed from the steady-state formulae. The non-lagging camber force part $F_{y,NL}$ is assumed to act directly on the wheel rim. We may approximate the non-lagging force part by a linear relation with γ using the camber thrust stiffness $C_{F\gamma}$, the non-lagging fraction ε_{NL}, and the weighting function $G_{y\kappa}$ to take care of the presence of a fore and aft force F_x:

$$F_{y,NL} = G_{yk}\varepsilon_{NL}C_{Fy}y \tag{17.201}$$

It should be noted that ε_{NL} changes with wheel load. We may refer to [1] for a more suitable and general approach to account for the non-lagging part.

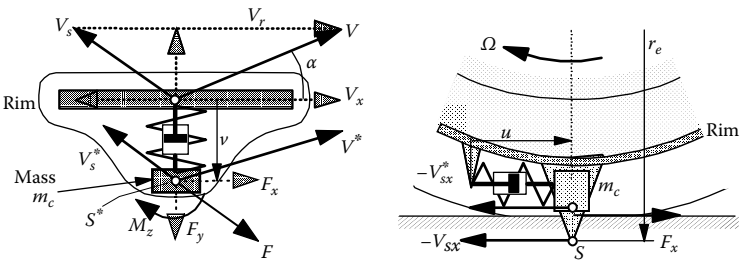


FIGURE 17.27 Enhanced transient tire model in top and side view showing carcass compliances and contact patch mass.

To enable calculations near or at standstill, we may add additional first-order differential equations with relaxation length σ_c. From these equations, the transient slip quantities result that act as input in the steady-state slip force formulae. We have

$$\sigma_c \frac{d\alpha'}{dt} + |V_x|\alpha' = -V_{sy}^* \tag{17.202}$$

$$\sigma_c \frac{d\kappa'}{dt} + |V_x|\kappa' = -V_{sx}^* \tag{17.203}$$

The contact relaxation length σ_c may be given a small but finite value, for instance, equal to half the contact length, a_o, at nominal load. This corresponds to theoretical findings of Section 17.6. Equations 17.202 and 17.203 do not respond to load changes. For this, we rely on the effect of the carcass compliance (in conjunction with the load-dependent cornering stiffness) that gives adequate results. Effectively, the resulting lateral compliance of the standing tire is

$$\frac{1}{C_{Fy}} = \frac{1}{c_{cy}} + \frac{\sigma_c}{C_{F\alpha}} \tag{17.204}$$

and the effective resulting tire relaxation length (for side slip response) is

$$\sigma_\alpha = \frac{C_{F\alpha}}{C_{Fy}} = \frac{C_{F\alpha}}{c_{cy}} + \sigma_c \tag{17.205}$$

This equation may, in fact, be employed to assess the lateral carcass stiffness at ground level c_{cy}. From the measured tire relaxation length σ_α and cornering stiffness $C_{F\alpha}$, the lateral stiffness of the standing tire C_{Fy} follows. By taking σ_c equal to half the contact length a, the carcass lateral stiffness c_{cy} can be determined. The relaxation length for load variations remains equal to

$$\sigma_{\alpha,Fz} = \frac{C_{F\alpha}}{c_{cy}} \tag{17.206}$$

which is a little smaller than σ_α. Although this results from practical modeling considerations, we may in fact come close to the measurement results (cf. ([1], Figure 7.10)).

If one is not interested to include the ability to simulate near or at zero forward speed, the contact relaxation length σ_c may be disregarded and taken equal to zero. In Section 17.6, the contact relaxation length σ_c forms an essential element in the model developed for short-wavelength behavior.

The deflection rates needed in Equations 17.199 and 17.200 are equal to the difference in slip velocities:

$$\dot{u} = V_{sx}^* - V_{sx} \tag{17.207}$$

$$\dot{v} = V_{sy}^* - V_{sy} \tag{17.208}$$

As has been mentioned before, the wheel slip velocity V_s with components $V_{sx,y}$ is defined as the horizontal velocity of the slip point S that is thought to be attached to the wheel rim at a distance r_e, the effective rolling radius, below the wheel center in the wheel center plane:

$$V_{sx} = V_x - r_e\Omega \tag{17.209}$$

$$V_{sy} = V_y - r\dot{\gamma} \tag{17.210}$$

where $V_{x,y}$ denote the horizontal (parallel to road plane) components of the wheel center velocity. In Equation 17.210, the camber angle is assumed to be small and r_e is replaced by the loaded radius r.

The forces acting on the wheel rim, finally, become

$$F_{xa} = k_{cx}\dot{u} + c_{cx}u \tag{17.211}$$

$$F_{ya} = k_{cy}\dot{v} + c_{cy}v + F_{y,NL} \tag{17.212}$$

The aligning torque becomes

$$M'_z = -t_\alpha F_y \tag{17.213}$$

$$M_{zr} = M_{zr}(\gamma, \alpha', \kappa', F_z) \tag{17.214}$$

$$M_z = M'_z + M_{zr} + s \cdot F_x + M_{z,gyr} \tag{17.215}$$

The last equation forms an extension of Equation 17.158 through the introduction of the gyroscopic couple that follows from Equation 17.196.

To illustrate the performance of the model, the calculated side force response to successive step changes in camber angle, slip angle, and brake slip has been presented in Figure 17.28. The contact patch relaxation length σ_c has not been used in this example. In [1], the model is applied in the problems of controlled braking on an uneven road and on starting from standstill where in the latter case σ_c is employed. The less demanding case of the response of side force and aligning torque to successive steps in pure side slip, shown in Figure 17.29, was calculated with the restricted fully

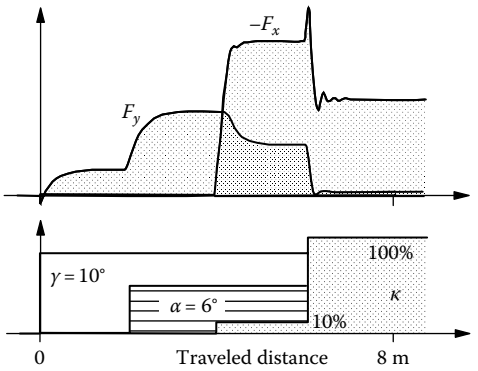


FIGURE 17.28 Side force and brake force response to step changes in combined slip, computed with the enhanced transient tire model of the present section.

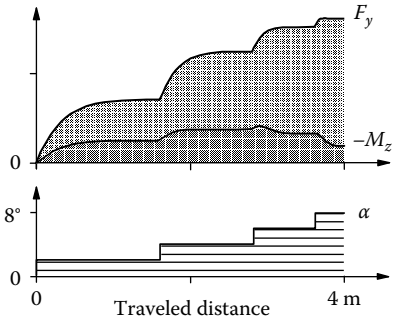


FIGURE 17.29 Side force and aligning torque response to successive steps in side slip as computed with the restricted transient fully nonlinear model, Equation 17.192.

nonlinear model according to Equation 17.192. Very similar results are found when the enhanced model is used, that is, with the inclusion of the contact mass m_c.

17.6 SHORT-WAVELENGTH INTERMEDIATE FREQUENCY TIRE MODEL

The present section briefly describes a model that is able to cover situations where the wavelength is relatively short (>20 cm and possibly even shorter for modeling road obstacle enveloping properties) and the frequency is relatively high (ca. <60–80 Hz) while the level of slip may be high. Situations in which combined slip occurs can be handled, and the Magic Formula model can be used as the basis for the nonlinear force and moment description. As a result, a continuous transition from time-varying slip situations to steady-state conditions is realized. The original model development was restricted to the more important responses to variations of longitudinal and side slip. Also, the possibility to traverse distinct road irregularities (cleats) was included in the tire model. The model is based on the work of Zegelaar [19] and Maurice [20] conducted at the Delft University of Technology and supported by TNO Automotive and a consortium of industries. The model is referred to as the *SWIFT* model (corresponding to the title of the present section) or in the context of Delft tire: *MF-Swift*. Subsequent developments of the model made it possible to also include variations of wheel camber and turn slip. Moreover, more general road surfaces with 3-D unevennesses can be handled. The reader is referred to [1,21] for detailed information.

The crucial step that was taken to reach further than one can by using the string model is the separation of modeling the carcass and the contact patch. In this way, a much more versatile model can be established that correctly describes slip properties at short wavelengths and at high levels of slip. The model achieved can be seen as a further development of the enhanced single-contact-point model of the preceding section, Figure 17.27. In Figure 17.30, the model structure has been depicted. Five elements of the model structure can be distinguished: (1) The inertia of the belt that has been taken into account to properly describe the dynamics of the tire. The restriction to frequencies of about 60–80 Hz allows the belt to be considered as a rigid circular ring. (2) The so-called residual stiffnesses that have been introduced between contact patch and ring to ensure that the total static tire stiffnesses in vertical, longitudinal, lateral, and yaw directions are correct. The total tire model compliance is made up of the carcass compliance, the residual compliance (in reality a part of the total carcass compliance), and the tread compliance. (3) The brush model that represents the contact patch featuring horizontal tread element compliance and partial sliding. On the basis of this model, the effects of the finite length and width of the footprint are approximately included. This element of

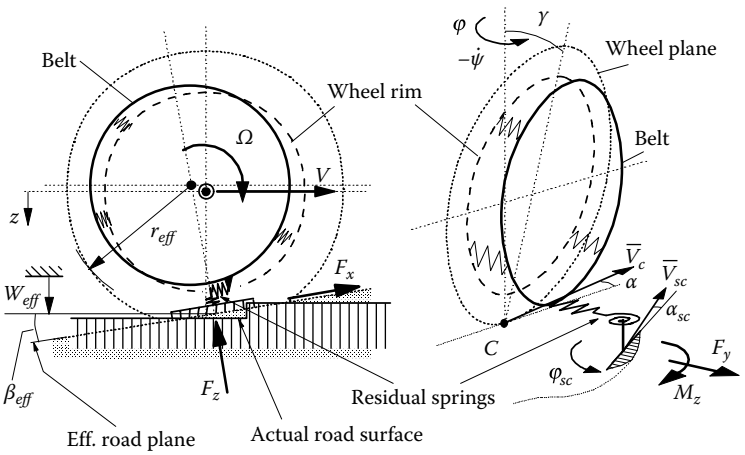


FIGURE 17.30 General configuration of the Swift model featuring rigid belt ring, residual stiffnesses, contact patch slip model, and effective road inputs.

the model is the most complex part and accomplishes the reduction of the allowed input wavelength to ca. 10 cm. (4) Effective road inputs to enable the simulation of the tire moving over an uneven road surface with the enveloping behavior of the tire properly represented. The actual 3-D profile of the road is replaced by a set of four effective inputs: the effective height, the effective forward and transverse slopes of the road plane, and the effective forward road curvature that is largely responsible for the variation of the tire's effective rolling radius. (5) The Magic Formula tire model to describe the nonlinear slip force and moment properties.

Similar but more physically oriented models have been developed. A notable example is the *BRIT* model of Gipser. He employs a brush-ring model featuring tread elements distributed over a finite contact patch with realistic pressure distributions and demonstrates the use of the model with simulations of the tire traversing a sinusoidal road surface also at a slip angle and at braking (cf. Gipser et al. [4]). Other models using discrete and finite elements, *Ftire* and *RMod-K*, have been developed by Gipser [5] and Oertel and Fandre [7], respectively. In [22] Lugner et al. give a comprehensive comparative discussion on the three dynamic models *MF-Swift*, *Ftire*, and *RMod-K*.

In the ensuing discussion, firstly, the slip model of the contact patch covering small wavelengths and large slip will be dealt with. Secondly, the model for the description of the dynamic behavior of the rigid belt ring with residual compliance will be added. Thirdly, the feature of the model that takes care of running over uneven roads will be addressed. Full-scale tire test results will demonstrate the validity of the model.

17.6.1 Contact Model Non-Steady-State Behavior

Theoretical analysis (cf. [1]) shows that the transient slip behavior of the brush model can be approximated by three first-order differential equations. They read for the transient longitudinal slip k'

$$\sigma_c \frac{d\kappa'}{dt} + |V_x|\kappa' = -V_{sxc} \tag{17.216}$$

the transient side slip α'

$$\sigma_c \frac{d\alpha'}{dt} + |V_x|\alpha' = -V_{syc} - |V_x|\psi'^{r}_{st} \tag{17.217}$$

and the transient side slip for the pneumatic trail α'_t

$$\sigma_2 \frac{d\alpha'_t}{dt} + |V_x|\alpha'_t = |V_x|\alpha' \tag{17.218}$$

Additional four first-order differential equations take care of the response to varying turn slip and camber. The forward speed is denoted by V_x. The inputs are the slip velocities of the contact patch, V_{sxc}, V_{syc}, and $d\psi_{sc}/dt$. The outputs are the transient slip quantities. The last term in (17.217) (static carcass yaw distortion in feedback loop) is needed to enable the use of actual tire side slip characteristics (17.220).

The relaxation lengths decrease with the magnitude of total slip, due to decreasing zone of adhesion in the contact patch. At vanishing slip we have $\sigma_c = a$ (half the contact length) and $\sigma_2 = t$ (the pneumatic trail). To assess the forces and moment, the transient slip quantities are subsequently used as input into the steady-state force and pneumatic trail formulae. At pure slip conditions we would have

$$F_x = F_x(\kappa', \alpha', \ldots) \tag{17.219}$$

$$F_y = F_y(\kappa', \alpha', \ldots.) \tag{17.220}$$

$$M_z = -t(\kappa', \alpha'_t \ldots) \cdot F_y + M_{zr}(\kappa', \alpha', \ldots) + (s + y^r_{st}) \cdot F_x + C_{\Delta M} \cdot (\alpha' - \alpha'_t) \tag{17.221}$$

The last term with $C_{\Delta M}$ in (17.221) represents the small extra moment, which is kept in linear form. This term is needed to model the situation where due to the differences in phase, the side force may become zero while the moment is still nonzero. The coefficient $C_{\Delta M}$ also decreases gradually to zero with increasing magnitude of the total slip. At vanishing slip this quantity is equal to the aligning stiffness of the contact patch model (the brush model). M_{zr} represents the moment due to spin (turning and camber). The term with F_x reflects the influence of the lateral shift of the line of action of F_x (cf. original Magic Formula model formulations).

At three different levels of side slip, α_o = 0, 0.08 and 0.16 rad, a comparison of the earlier approximate simulation model (for the brush model with (17.219) through (17.221) replaced by its analytically established steady-state characteristics and the last term of (17.217) removed) with the analytical solution of the brush physical model has been conducted in terms of the path frequency response functions. Figure 17.31 shows the results for the case that the contact patch model is attached to a flexible carcass. The general conclusion is that at least for wavelengths larger than ca.10 cm, the correspondence can be judged to be very good. The upper pair of diagrams that refer to the side force response shows the sideways "shift" of the phase lag curves that is caused by the drop in relaxation length with increasing average slip angle. The diagram for the aligning torque response clearly shows the transition from second- to first-order behavior when the average slip angle changes from zero to larger values.

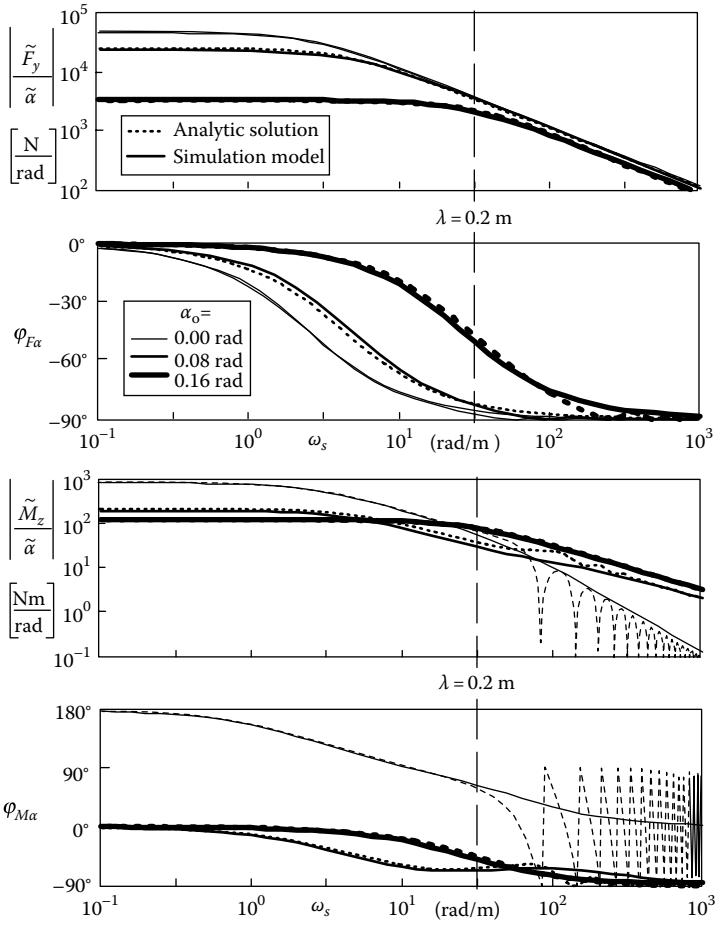


FIGURE 17.31 Frequency response functions of linearized system including carcass flexibility at different average slip angle levels according to the analytical solution and to the approximate simulation model. The path frequency at a wavelength of the input wheel plane motion equal to 20 cm has been indicated.

The model may be expanded also to cover the responses to changes in camber and turn slip. The latter is important especially for the simulation of parking maneuvers. We refer to the original text [1] for detailed information.

The first problem that was encountered is the fact that in the model described earlier where the contact patch is represented by the brush model, the steady-state characteristics employed (17.220, 17.221) belong to the brush model and not to the total model including yaw torsional compliance of the carcass. It is preferred to use functions in terms of the slip values of the wheel, since these functions (possibly Magic Formulae) are directly available from full-scale experiments performed on the road or on flat track or drum test rigs.

An obvious solution is to model the contact patch characteristics with the Magic Formulae. These deviate from those assessed for the complete tire because the contact patch "sees" a slip angle that differs (is smaller) from that of the wheel plane. A set of adapted *MF* parameters may be established off-line for the contact patch, or an iteration loop may be included to achieve the correct steady-state behavior of the total model. A simpler, more practical way has been found, however, which allows the use of the Magic Formula characteristics as obtained from tire slip tests. This is accomplished by adding the last term of Equation 17.217 (the static yaw distortion of the carcass) (cf. [1]).

17.6.2 Residual Stiffnesses and Belt Dynamics

In the preceding part, the contact patch model has been used in connection with a flexible carcass. In the present section, the inertia of the belt is introduced. Since we restrict the application of the model to frequencies lower than ca. 60–80 Hz, the belt may be approximated as a rigid ring that is attached to the wheel rim through flexible sidewalls. To ensure that the total static tire stiffness remains unchanged, residual springs have been introduced between contact patch and belt. The residual springs defined in the model have been indicated in Figure 17.32.

The horizontal dynamics of the contact patch are handled with a set of six first-order differential equations, while the six degrees of freedom belt motion are modeled with two sets of three second-order differential equations.

In the study of Zegelaar [19], also reported in [1], important experimental observations have been made regarding constitutive relationships of the tire such as contact area dimensions, static and dynamic vertical stiffness and characteristics at different speeds of rolling, static longitudinal stiffness of the standing tire, tire radius growth with speed, rolling resistance, effective rolling radius, and rolling resistance couple. Model parameter values are assessed through full-scale tire dynamic experiments carried out on different types of test facilities. Many of the parameters are obtained by minimizing the difference between measured and calculated (complex) response functions.

The calculated vibrational rigid body modes at zero speed using the parameters as established from experiments carried out on the rotating drum have been depicted in Figures 17.33 and 17.34.

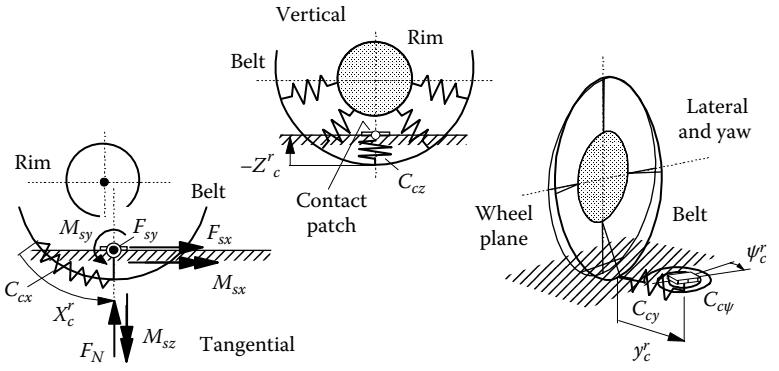


FIGURE 17.32 Deflections of the residual springs attaching contact patch to belt.

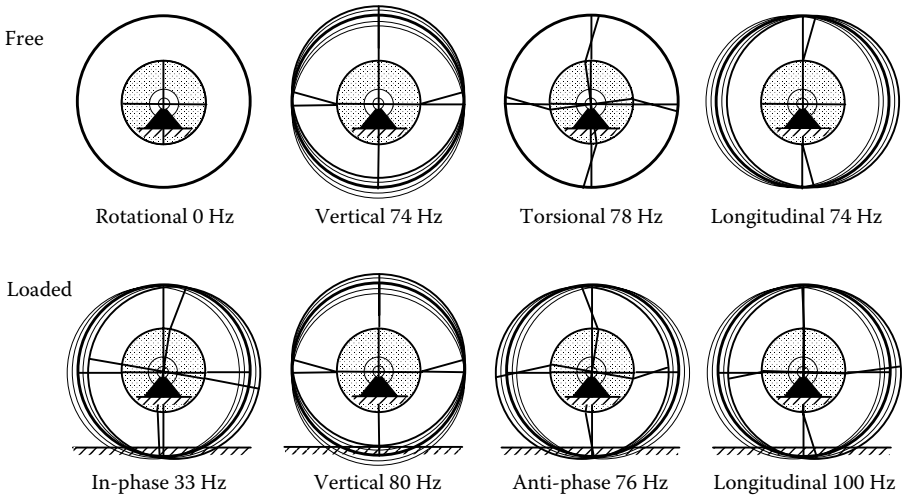


FIGURE 17.33 Calculated in-plane vibrational modes of tire/wheel system with axle fixed and tire free or loaded on the drum surface with vertical load F_N = 4000 N and at zero speed.

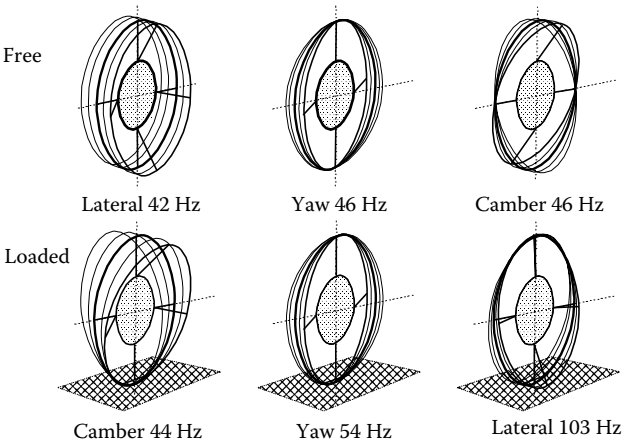


FIGURE 17.34 Calculated out-of-plane vibrational modes of tire/wheel system with axle fixed and tire free or loaded on the drum surface with vertical load F_N = 4000 N and zero speed.

We have four in-plane degrees of freedom of belt ring and wheel rim (two translational and two rotational) and three out-of-plane degrees of freedom (lateral, yaw, and camber). As a consequence, we can distinguish four in-plane rigid body modes and three out-of-plane rigid body modes. The mode shapes may change considerably when the tire is making contact to the drum surface. The free rotation (0 Hz) mode changes into a mode with the belt and rim rotating in phase with respect to each other. The lateral and camber modes form combinations: one low-frequency mode with a low axis of camber oscillations and one high-frequency mode with a rotation axis closer to the top of the tire and relatively large lateral deflections in the contact zone.

As an example, Figure 17.35 presents the out-of-plane frequency response functions of the side force and the aligning torque amplitudes to yaw oscillations compared with test results as obtained on the yaw oscillation test stand up to ca. 60 Hz. To demonstrate the in-plane performance of the model, simulations and experiments conducted on the dynamic brake/cleat drum

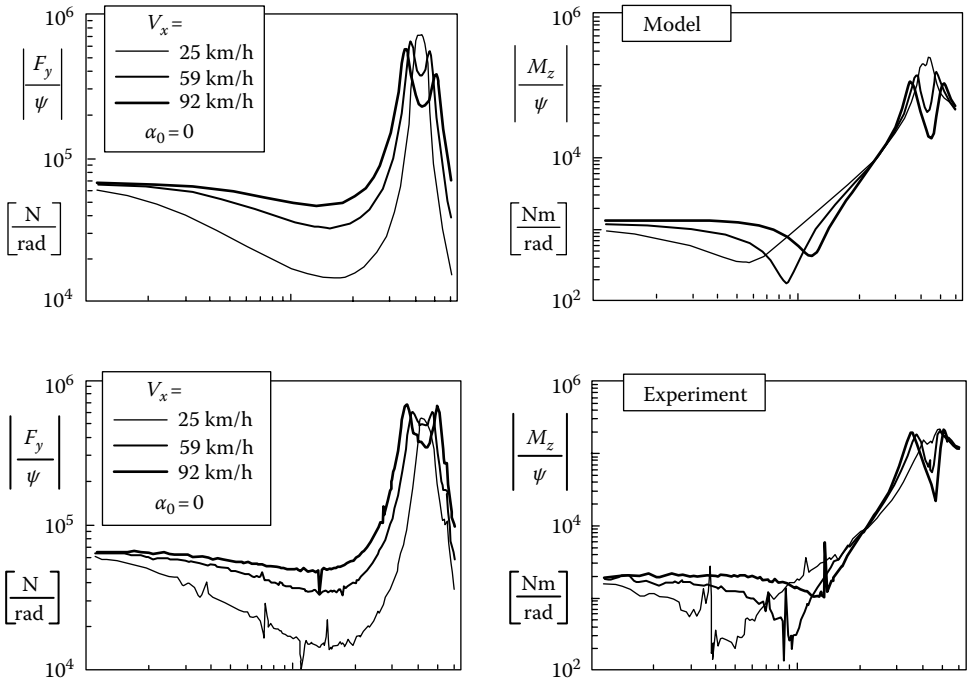


FIGURE 17.35 Calculated (top) and measured out-of-plane frequency response functions of the side force and aligning torque amplitude to steer angle variations at zero average slip angle for three values of forward speed and normal load $F_N = 4000$ N. (From Maurice, J.P., Short wavelength and dynamic tire behaviour under lateral and combined slip conditions, Dissertation, TU Delft, Delft, the Netherlands, 2000.)

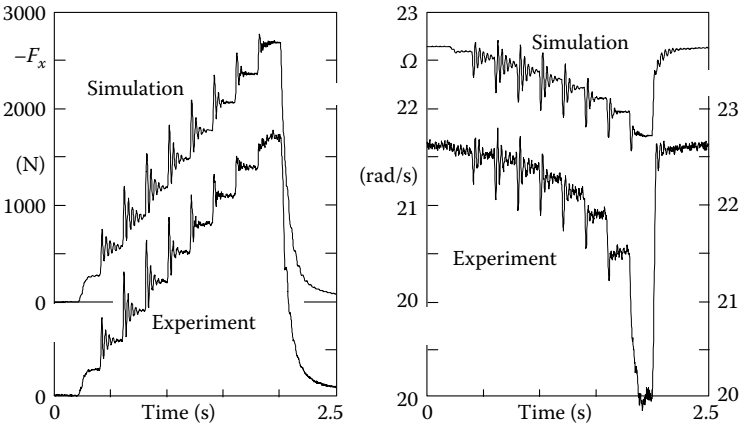


FIGURE 17.36 Brake force response to successive step increments of brake pressure. The diagrams depict the variation of force and wheel speed of revolution with time. Apparently, the actual friction coefficient is lower than assumed in the model. (From Zegelaar, P.W.A., The dynamic response of tires to brake torque variations and road unevennesses, Dissertation, TU Delft, Delft, the Netherlands, 1998.)

test facility have been carried out pertaining to successive stepwise increases in brake pressure. The response of the longitudinal force and the associated wheel speed has been presented in Figure 17.36. Apparently, the rigid ring model provided with the short-wavelength transient slip model and residual compliances is very well capable of describing the dynamic tire behavior in the frequency range up to about 60 Hz.

17.6.3 Dynamic Tire Response to Short Road Unevennesses

The actual road surface profile over which the tire rolls may contain spectral components showing relatively short wavelengths. If the wavelength is smaller than two to three times the contact length, a geometric filtering of the profile becomes necessary if the tire model employed is assumed to contact the road in a single point. For the *SWIFT* model, a special filter has been developed that takes care of the envelopment properties of the tire and the variation in effective rolling radius that occurs when the tire rolls over a short obstacle. The envelopment of an obstacle takes place in the contact zone. It is assumed that local dynamic effects can be neglected. The changing conditions that arise for the tire when quasi-statically traversing an obstacle are measured and modeled and subsequently used as effective road plane inputs for the tire model at higher speeds. The belt inertia takes care of the dynamic effects. To arrive at a geometrically filtered road profile description, Bandel [23] introduces the effective road plane. The effective plane height and forward slope variation may be established by conducting an experiment where the wheel is rolled at a very low velocity over an uneven road surface at constant vertical load and the forces and vertical displacement are measured. According to the definition, the resulting force (with the rolling resistance force subtracted) that acts upon the wheel axle is directed perpendicularly to the effective road plane. The effective height variation, w, is taken equal to the vertical axle displacement. The effective forward slope $\tan\beta_y$ is found by dividing the longitudinal horizontal force (after having subtracted the relatively small rolling resistance force) by the vertical force. In the model, the effective road height and slope can be assessed by using basic curves in connection with a two-point follower with shift length l_s (Zegelaar [19]) or by moving two cams with an oval shape (Schmeitz, cf. [1 (3d ed.), 21,24]) over the actual road surface following each other at a distance equal to the length l_s (slightly smaller than the contact length). The height of the midpoint of the connecting line of the lowest points of the cams and the inclination of this line represent the effective road height and effective forward slope. Figure 17.37 illustrates this "tandem cam" configuration. Fitting the tandem cam parameters is accomplished by assessing the best approximation of very low-speed responses of a tire running over steps or strips of different heights at a number of constant vertical loads. Once the parameters have been established, the cam dimensions can be kept fixed and independent of step height, obstacle shape, or vertical load. Inspection shows that the shape of the lower part of the ellipse turns out to be practically identical to the contour of the tire in side view just in front of the contact zone up to the height of the highest step considered in the fitting process. The tandem base length l_s does depend on the vertical load. The tandem cam configuration can be employed to simulate rolling over single cleats or arbitrary irregular, normally encountered, road surfaces.

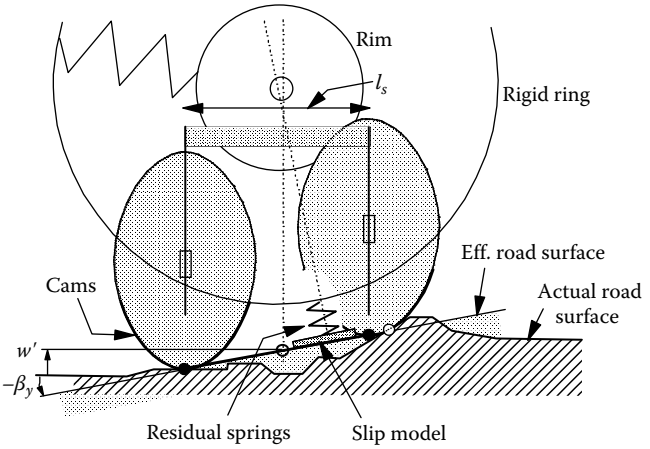


FIGURE 17.37 The "tandem cam" configuration of Schmeitz, generating the effective road height and slope, attached to the ring and slip model.

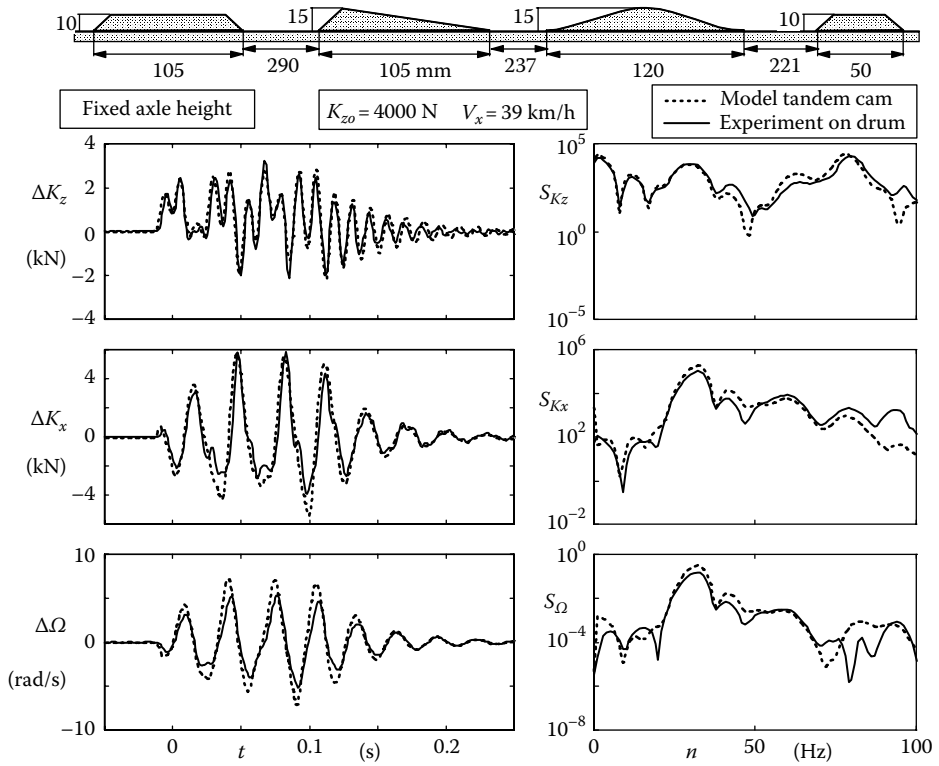


FIGURE 17.38 Running over a series of cleats mounted on the drum surface, at constant axle height. Measured and computed variations of vertical and longitudinal wheel spindle forces K_z and K_y and wheel angular speed Ω versus time and as power spectral densities S. The center of the first (left-hand) cleat passes under the wheel center at $t = 0$. After t = ca. 0.12 s the surface is undisturbed.

Analysis shows that the variation of the effective rolling radius with respect to the initial condition, where the effective slope β_y and slope rate of change (curvature) $d\beta_y/ds$ are zero and the radial tire deflection $\rho_z = \rho_{zo}$, becomes

$$\tilde{r}_e = -\eta\tilde{\rho}_z - r_{eo}(1-\cos\beta_y) + \rho_z r_{eo}\frac{d\beta_y}{ds} \tag{17.222}$$

On uneven road surfaces, the last term constitutes by far the most important contribution to the effective rolling radius variation and thus to the wheel rotational acceleration that can only be brought about by a variation in the longitudinal contact force F_x. This force also often outweighs the part of the horizontal longitudinal force that directly results from the forward slope itself.

In Figure 17.38, an example is presented where a 205/60 R15 tire rolls at constant axle height over a drum surface provided with a number of differently shaped obstacles. Comparison of measured and computed vertical and horizontal (longitudinal) wheel spindle forces K_z and K_x and wheel speed of revolution Ω varying with time and as power spectra S shows satisfactory agreement.

Road camber variations that occur, for example, when the tire rolls over an oblique cleat, can be handled by taking two parallel tandems (track width taken equal to the width of the contact patch) and assessing the average transverse effective slope tan β_x (cf. Figure 17.39 showing the double-track "road feeler" moving over a step-shaped road unevenness with angle θ). For short-wavelength (<0.2 m) road camber variations that, for example, occur with obstacles like oblique steps or strips, additional intermediate cams (say three additional cams along each of the four edges of the "road feeler") are to be inserted with slopes averaged to get sufficiently accurate results (cf. [21,24]).

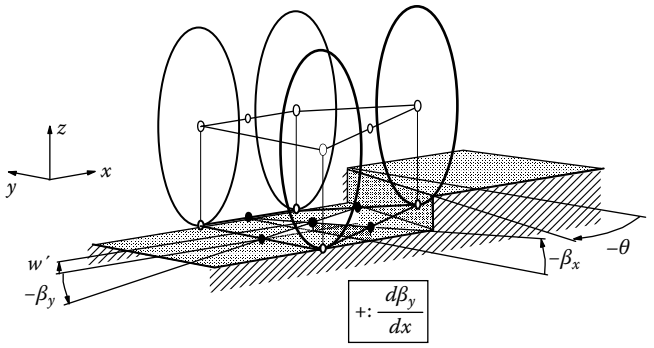


FIGURE 17.39 The double-track tandem cam road feeler for 3-D road unevennesses, moving over oblique step. Four effective road inputs.

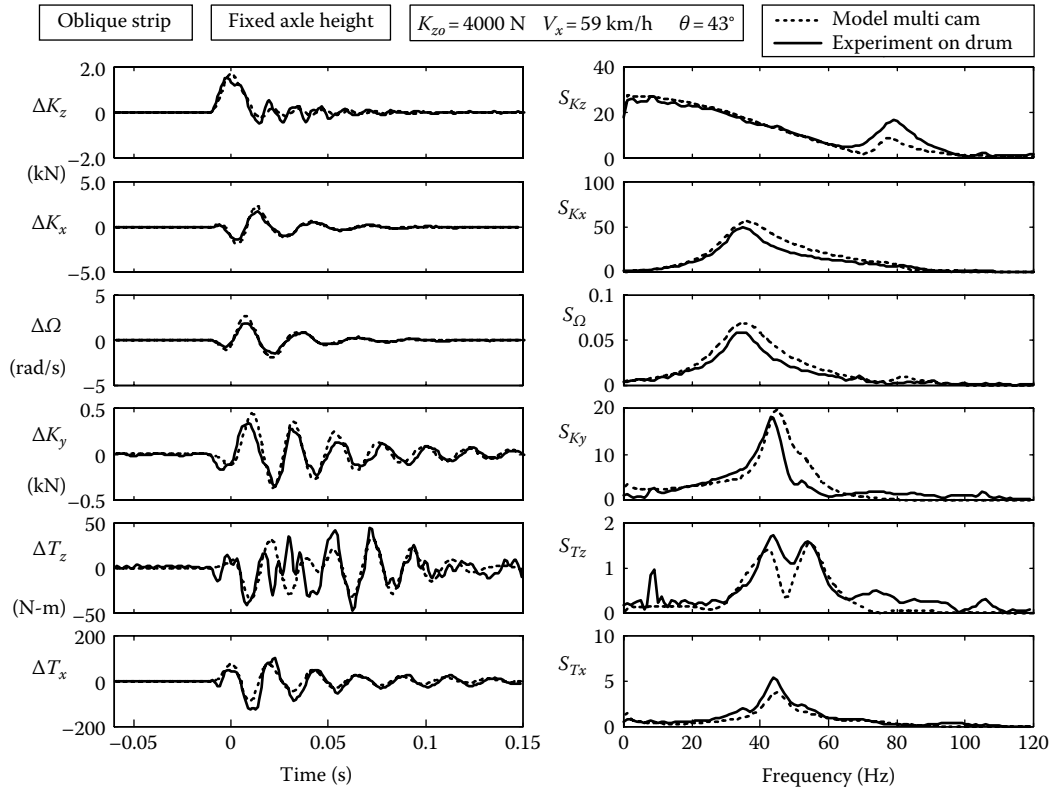


FIGURE 17.40 Running over an oblique cleat (10 × 50 mm strip) mounted on the drum surface at an angle of 43°, at constant axle height. Measured and computed variations of wheel spindle forces K and moments T and wheel angular speed Ω versus time and as power spectral densities S.

In total we have now four effective road input quantities: (1) the effective height w, (2) the effective forward road slope angle β_y, (3) the effective transverse road slope angle β_x, and (4) the effective forward road curvature $d\beta_y/dx$.

In Figure 17.40, example results are shown of experimental and model simulation results of the tire moving at constant axle height at a speed of 59 km/h over an oblique strip (height 10 mm, width 50 mm) mounted at an angle of 43° with respect to the drum axis (wheel slip angle and camber angle are zero). The variations of the wheel spindle forces $K_{x,y,z}$ and moments $T_{x,z}$ and the wheel angular velocity Ω are presented as functions of time and as power spectral densities versus frequency.

Again, it is seen that a reasonable or good correspondence can be achieved. The various tire/wheel resonance frequencies can be observed to show up vertical mode at 78 Hz (K_z), rotational mode at 35 Hz (K_x and Ω), camber mode at 46 Hz (K_y, T_z, and T_x), and yaw mode at 54 Hz (T_z).

17.7 TIRE STEADY-STATE AND DYNAMIC TEST FACILITIES

At various automotive and tire industrial companies and at a number of universities and institutes, test facilities are available for performing full-scale tire measurements to assess the tire force and moment generation properties. The test installation may be built on a truck or trailer that is equipped with a special wheel suspension and guidance system to which a measuring hub is attached. Typically, with such an over the road testing equipment, moderate speeds up to ca. 120 km/h can be reached. The frequencies with which the steer angle can be varied are relatively low. The camber angle is changed mechanically or through a hydraulic cylinder. The vertical load may be set at a desired average level; the load variations caused by road unevennesses are filtered out. Commonly, the longitudinal slip results from the controlled application of the brake pressure. In very few cases, the wheel angular velocity is controlled through a hydraulic motor that acts relative to the vehicle's road wheel speed of revolution. In such a way, the test wheel drive and brake slip can be varied in a controlled manner. In some devices, water can be sprayed in front of the test wheel to create wet road conditions. The measuring hub contains strain gauge or piezoelectric force measuring elements. With these test facilities usually measurements are conducted at quasi-steady-state conditions. Typically, a side slip or brake slip sweep may be imposed at a low rate. After processing, correcting, and averaging the signals, the steady-state force and moment slip characteristics are obtained.

The large indoor test stands usually operate along similar lines. These rigs are based on an imitated road surface provided by the surface of a drum or by that of a flat track (endless belt). Somewhat higher frequencies of changing vertical axle position and yaw angle can usually be achieved (up to ca. 2–8 Hz). Drum test stands have been built with diameters ranging from 2 to ca. 4–5 m. With the larger drums the tire usually runs or can also be run on the inner surface. This configuration makes it possible to mount realistic road surface segments and to maintain a layer of water on the inner surface, thereby enabling testing at wet or icy (and even snowy) conditions. External drums of 2.5–3 m diameter are more often encountered. Also, flat bed and flat plank test rigs are used. Typically, but not always, these machines operate at very low speed of travel of either the plank or the wheel axle. Also, turn table or swing arm devices exist that are constructed to measure turn slip properties.

On the drum test stand, a measuring tower or a special wheel guidance system with a force measurement platform or hub can be positioned. For dynamic higher-frequency tire tests, very stiff rigs are required (or very soft seismic systems). Special equipment is used that is often limited to a specific application to allow the system to become light and sufficiently rigid, that is, high first natural frequency. Devices exist dedicated to specific tasks such as systems with axle fixed (but adjustable in height) to assess tire nonuniformity, cleat response, or response to brake pressure variations; axle forced to perform only vertical axle oscillations to assess the vertical dynamic stiffness and the response of the longitudinal force; and only yaw angle variation of the axle to assess the tire dynamic steer response.

In Figure 17.41, a picture is shown of the Delft tire test trailer (cf. [25] cf. [1] 3d edn.) now owned by TNO, Helmond. The trailer has two measuring stations: one for passenger car-size tires limited to not so large camber angles (<30°) and, on the other side, a system specially designed for motorcycle tires that can handle also large camber angles (up to 55°). The test wheel can be subjected to a (varying) steer and camber angle and can be braked up to wheel lock. Water can be sprayed at a controlled rate in front of the test tire. The three forces and two moments (i.e., except the brake torque) are measured with the aid of measuring hubs: a large strain gauge-based and a compact piezoelectric-based system, respectively.

Figure 17.42 depicts the flat track machine of MTS. The upper structure, in which the steering axis is located, can be tilted about a line that forms the line of intersection of the wheel center plane and the belt that imitates the road surface. Through this configuration a pure camber or wheel inclination angle variation can be created (at zero steer angle). A very early

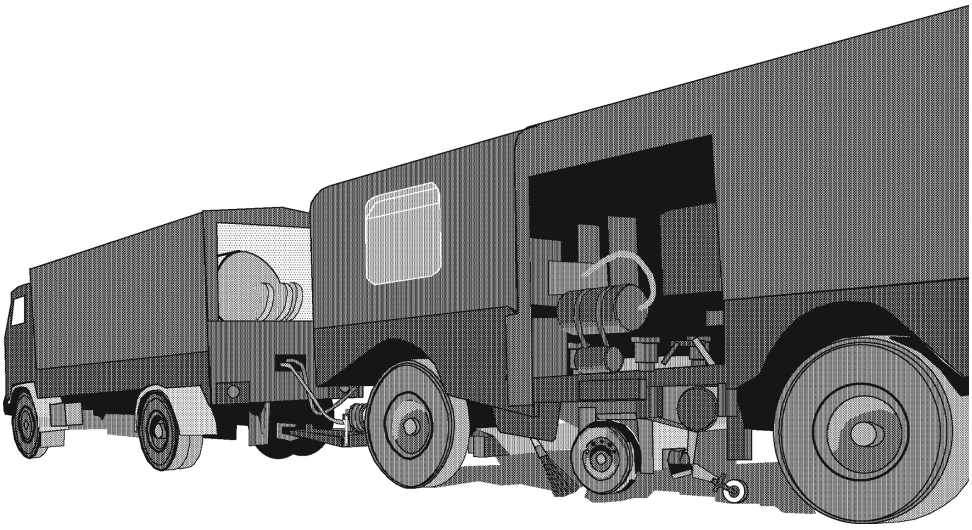


FIGURE 17.41 The Delft tire test trailer.

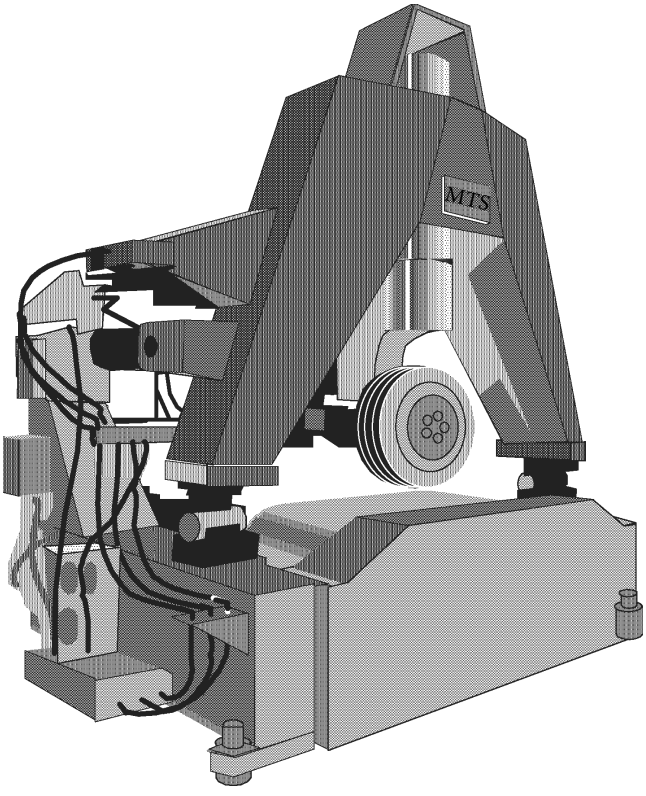


FIGURE 17.42 The MTS flat track test machine.

example of such a system is the large "Tread Mill" of (formerly) Calspan, Buffalo, that can also handle larger truck tires. The belt is supported by two drums and underneath the tire by a flat air or water-bearing surface. The lateral stabilization of the belt is accomplished by varying the yaw or the tilt angle of one of the drums. The test wheel can be driven and braked in a slip-controlled manner.

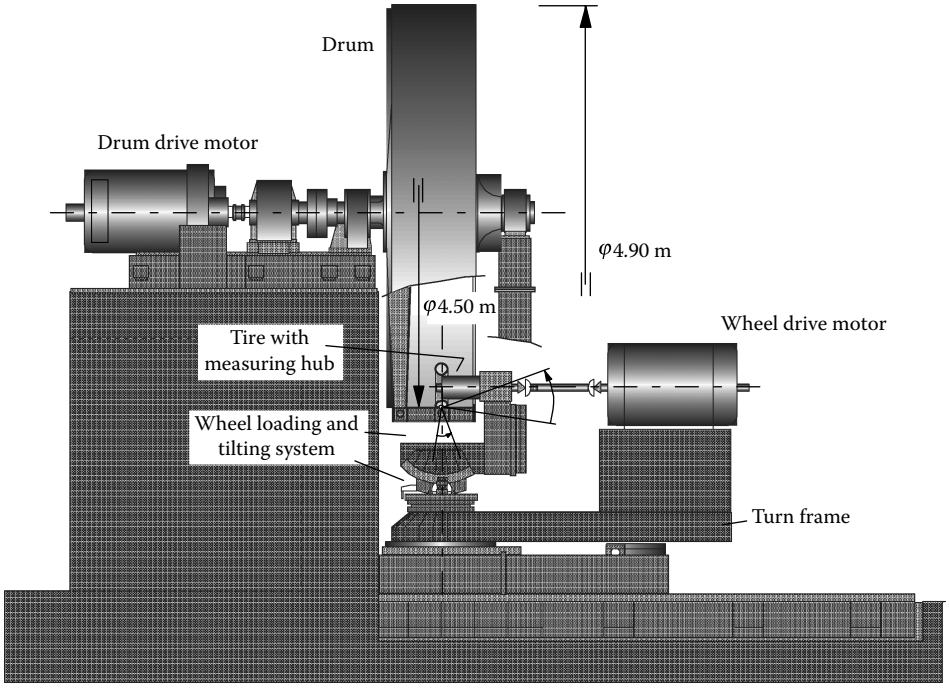


FIGURE 17.43 Internal drum test stand originally possessed and operated by Porsche.

Figures 17.43 and 17.44 show the layout of two internal drum test stands (cf. [26,27]). The drum of Figure 17.43 can also be used on its outer surface with a maximum speed of 250 km/h. These large rigs are operated at the Karlsruhe University of Technology. The arrangement of the wheel loading, tilting, and steering system enables center point steering and tilting about a line that touches the inner drum surface and lies in the wheel center plane. The order of changing the wheel attitude angles, first steering about the always vertical axis and then tilting about the new (still horizontal) x-axis, ensures that each of the angles, defined about and with respect to the vertical, remains unchanged when the other is varied. Of course, this principle also holds for the system of Figure 17.41 but not for the flat track machine of Figure 17.42. The configuration of Figure 17.44 exhibits a hydraulic wheel drive motor with which the wheel slip ratio can be controlled. Here, the force measuring unit rotates together with the wheel. This avoids the otherwise necessary constructional measure to suppress parasitic forces and cross talk such as the brake torque interaction with aligning torque that arises due to misalignment. With a stationary measuring hub, a double-Cardan-coupling drive shaft (including a length change compensation element) is an example of such a measure that ensures that practically only the drive/brake torque is transmitted to the wheel. Instead of a pair of these couplings, a set of two membranes (thin flexible discs) or other devices (see Figure 17.46) may be used. Accurate alignment remains necessary. With the UMTRI configuration (University of Michigan), cross talk has been prevented by positioning the brake system between the nonrotating, stationary measuring system and the tire. One of the drawbacks of rotating measuring systems is the fact that the sensitivity about the vertical axis (for the small aligning torque) cannot be chosen larger than the sensitivity about the horizontal longitudinal axis about which the moment may become relatively large.

Figure 17.45 shows the Delft flat plank machine (cf. [16,25]) now owned by the Eindhoven University of Technology. With this rig, accurate measurements can be conducted at a low speed of 2.3 cm/s. The plank has a maximum stroke of 7.5 m and can be tilted about the longitudinal center line on the test surface. Hereby, pure camber step responses can be established. The wheel axle, equipped with a measuring hub, can be steered and cambered, and the wheel can be braked. The vertical axle position or the vertical tire load can be adjusted. Cleats can be mounted on the plank surface. Typically, tire static

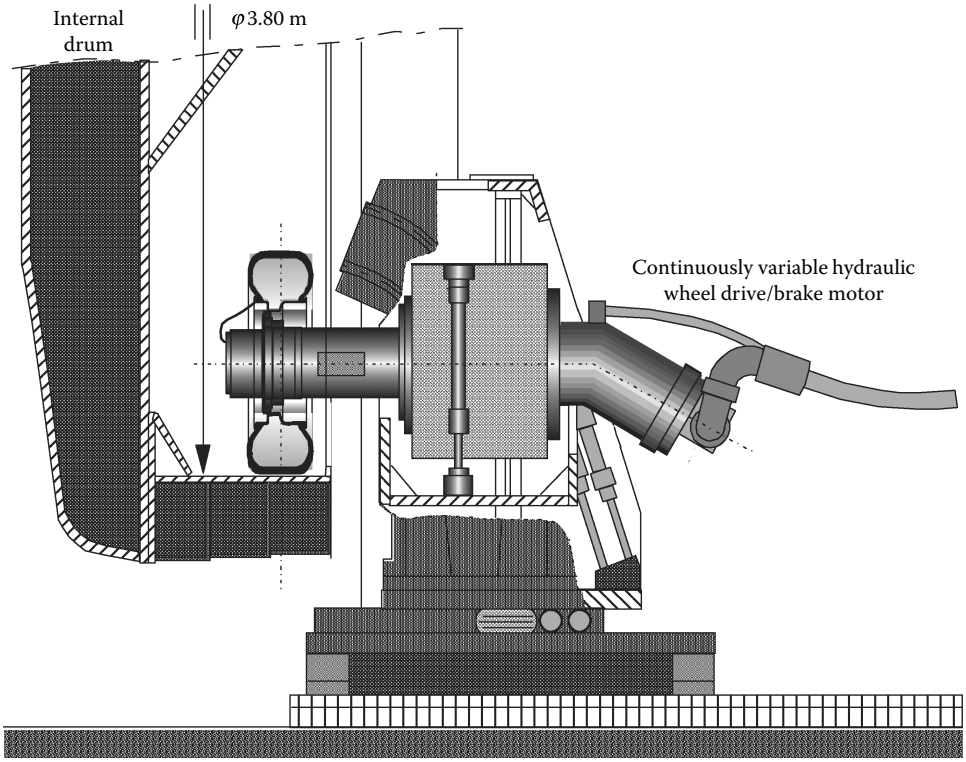


FIGURE 17.44 The Karlsruhe internal drum test stand.

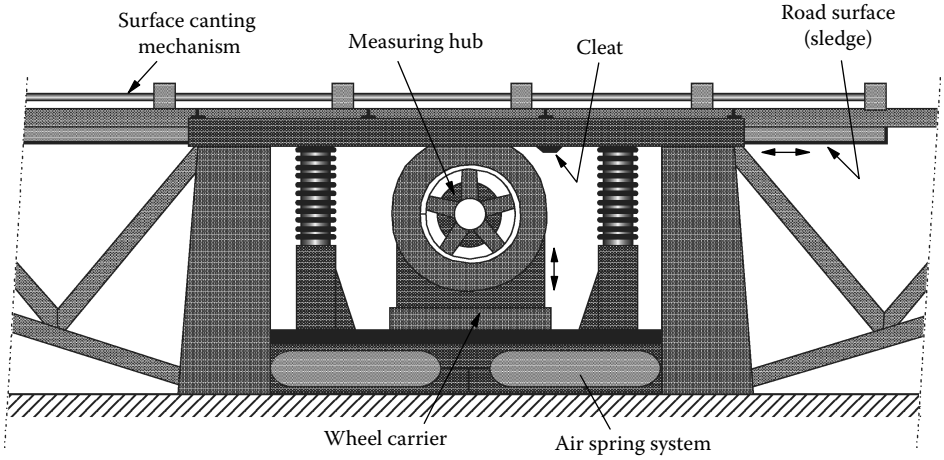


FIGURE 17.45 The flat plank machine with stroke of 7 m and road canting system.

stiffness tests (vertical, longitudinal, lateral, yaw, and camber), transient (step) side slip and camber tests (relaxation length), impulse turning, and cleat tests are performed on this machine.

The indoor drum test facility of the Delft University of Technology is based on two coupled drums with a diameter of 2.5 m that can run up to a maximum speed of 300 km/h. On top of one of the drums, a measuring tower (for low- or moderate-frequency yaw and brake experiments) can be installed (not shown). This rig [20,25] can be turned about a vertical axis that passes through the tire contact center and through the top of the drum surface. On the other drum, a rig can be mounted that is designed for measuring in-plane tire dynamics (Figure 17.46, [19]). Much care has been taken to make the

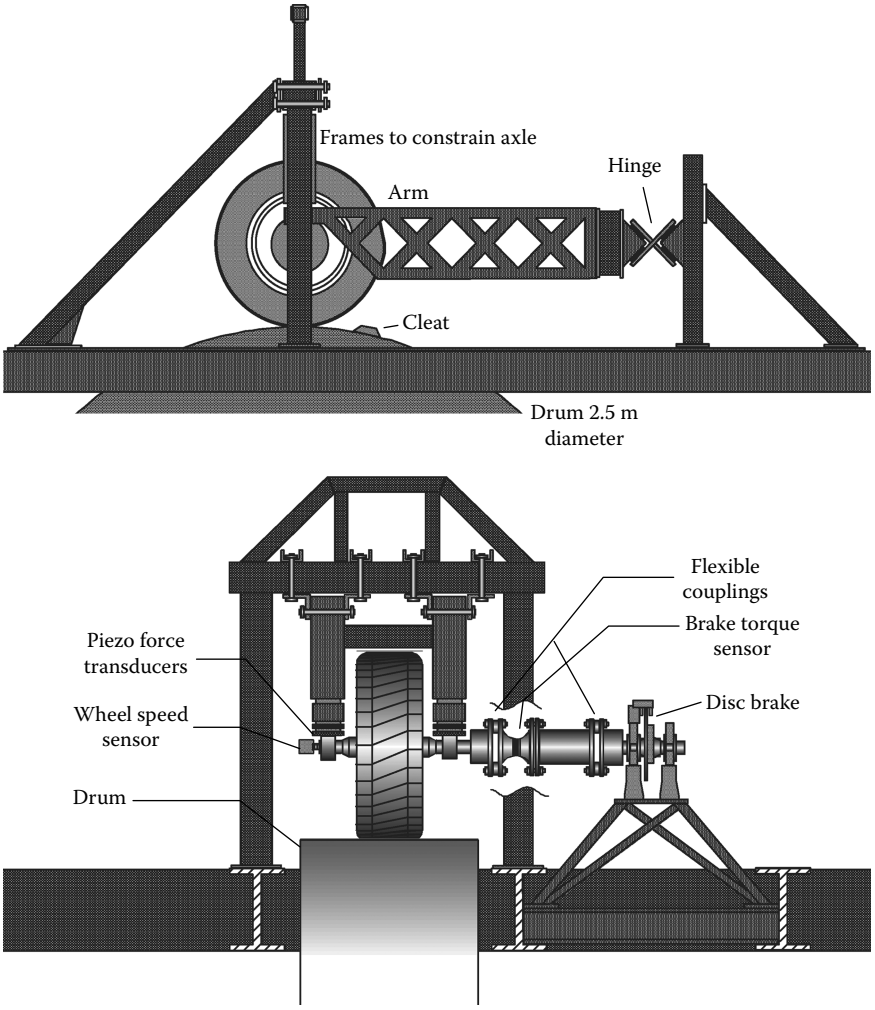


FIGURE 17.46 Dynamic brake and cleat test facility in side and front view. Tests may be conducted up to ca. 65 Hz enabling the assessment of tire/wheel in-plane inertia and stiffness parameters including residual stiffnesses and rigid modes of vibration.

rig sufficiently rigid. This resulted in a lowest natural frequency of just over 100 Hz that allows the use of test data up to ca. 70 Hz. In addition, to avoid force and moment cross talk, the brake shaft is connected with the wheel shaft through an intermediate shaft with two flexible couplings. These couplings represent an alternative solution for the double-Cardan coupling of Figure 17.43. They are flexible in all directions except about the axis of rotation. If properly aligned, this ensures that only the brake torque is transmitted to the wheel and that other parasitic forces and moments are largely suppressed. The brake torque is measured with strain gauges attached to the intermediate coupling shaft. A hydraulic servo system is used to control the brake pressure fluctuations (bandwidth up to ca. 60 Hz). Piezoelectric load cells placed on top of the bearings of the wheel shaft provide the signals from which the forces and moments that act on the wheel (except the brake torque) can be derived. In a different setup the machine is used for response measurements to vertical axle oscillations (<20 Hz).

On the other drum, rigs may be mounted for out-of-plane tire dynamic experiments. One is the trailing arm "pendulum" test stand [14,20] with at one end a vertical hinge and at the other the steering head (for adjusting the average slip angle) with piezoelectric measuring hub (cf. Figure 17.47). At that point the arm is excited laterally up to ca. 25 Hz by means of a hydraulic actuator. The wheel

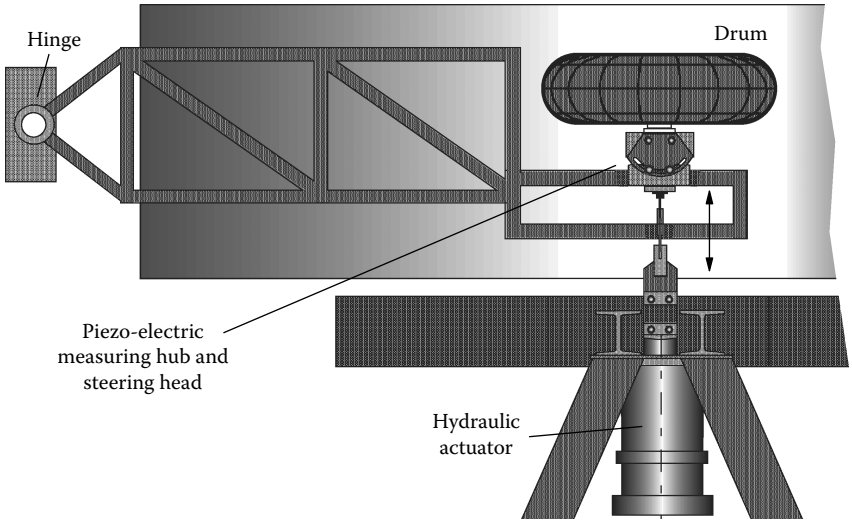


FIGURE 17.47 The trailing arm "pendulum" test rig exciting the tire almost purely laterally. Frequencies up to ca. 25 Hz adequate for assessing the tire relaxation length and gyroscopic couple parameter.

load is adjusted by tilting the vertical hinge slightly forward. The arm length is 1.65 m and the tire is subjected to an almost purely lateral slip variation. The rig is useful to assess the overall relaxation length and the gyroscopic couple parameter. In Figure 17.48, the so-called yaw oscillation test stand [20] is depicted that can be used for tests around an average steer angle that can be set at a value between −5° and +5°. The structure is light and very stiff. The two guiding members with flexible hinges intersect in the vertical virtual steering axis that is positioned in the wheel center plane (center point steering). A hydraulic actuator is mounted to generate the yaw vibration (typically random with a bandwidth of 65 Hz). The wheel axle is provided with a piezoelectric measuring hub. The tire is loaded by mechanically adjusting the axle height above the drum surface.

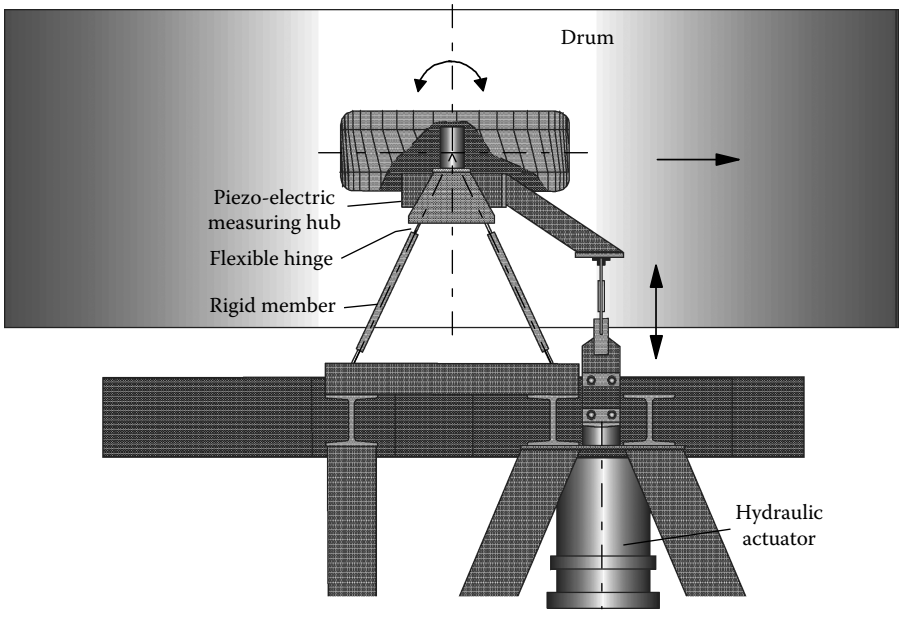
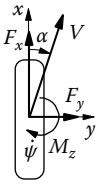
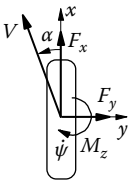
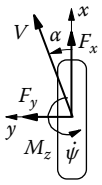
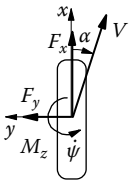
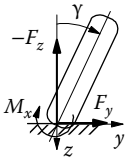
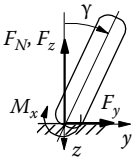
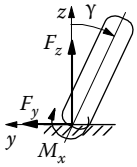
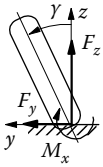
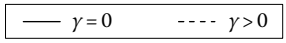
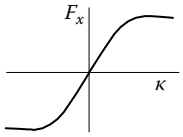
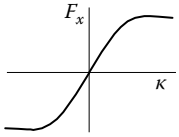
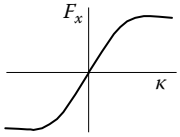
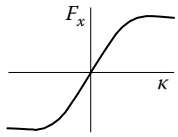
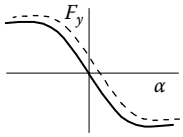
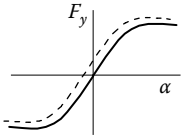
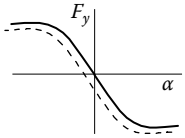
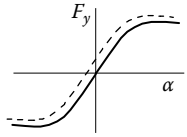
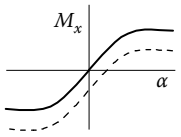
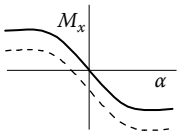
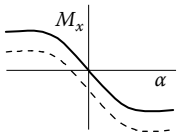
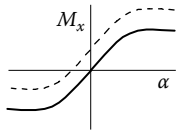
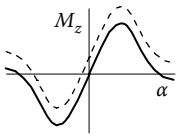
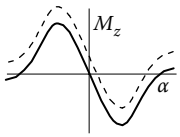
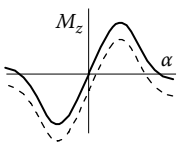
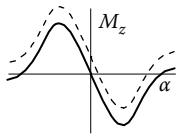


FIGURE 17.48 The yaw oscillation test rig featuring center point steering. Frequencies up to ca. 65 Hz enabling the assessment of tire out-of-plane inertia and stiffness parameters including residual stiffnesses and rigid modes.

17.A APPENDIX: SIGN CONVENTIONS FOR FORCE AND MOMENT AND WHEEL SLIP

($V_x > 0$)	SAE	Adapted SAE (Pacejka [1])	ISO	Adapted ISO (Besselink [28])
Side angle (top view)				
Inclination/ camber angle (rear view)				
Side slip	$\tan\alpha = \frac{V_{sy}}{V_x}$	$\tan\alpha = -\frac{V_{sy}}{V_x}$	$\tan\alpha = \frac{V_{sy}}{V_x}$	$\tan\alpha = -\frac{V_{sy}}{V_x}$
Longitudinal slip	$\kappa = -\frac{V_{sx}}{V_x}$	$\kappa = -\frac{V_{sx}}{V_x}$	$\kappa = -\frac{V_{sx}}{V_x}$	$\kappa = -\frac{V_{sx}}{V_x}$
Turn slip	Not defined	$\varphi = -\frac{\dot{\psi}}{V_x}$	Not defined	$\varphi = -\frac{\dot{\psi}}{V_x}$



Longitudinal force				
Side force				
Wheel load	$F_z < 0$	$> 0^*$	> 0	> 0
Overturning moment				
Rolling resistance moment	$M_y > 0$	> 0	< 0	< 0
Self-aligning moment				

*Except for Chapter 9 in [1] where $F_N = -F_z > 0$

REFERENCES

1. Pacejka, H.B., *Tyre and Vehicle Dynamics*. Butterworth and Heinemann (also SAE). Oxford, U.K., 2002, 2nd edn., 2006; 3ed., 2013.
2. Pacejka, H.B. and Sharp, R.S., Shear force development by pneumatic tyres in steady state conditions: A review of modelling aspects. *Vehicle System Dynamics*, 20, 121–176, 1991.
3. Pacejka, H.B. and Fancher, P.S., Hybrid simulation of shear force development of a tire experiencing longitudinal and lateral slip. In: *Proceedings of the XIV FISITA International Automotive Technology Congress*, London, U.K., 1972.
4. Gipser, M., Hofer, R., and Lugner, P., Dynamical tire forces response to road unevennesses. In: *Proceedings of the 2nd Colloquium on Tyre Models for Vehicle Analysis*, Eds. F. Böhm and H.P. Willumeit, Berlin, Germany, 1997, *Vehicle System Dynamics*, 27(Suppl.), 1996.
5. Gipser, M., Ftire, a new fast tire model for ride comfort simulations. In: *Proceedings of the International ADAMS User Conference*, Berlin, Germany, 1999.
6. Oertel, Ch., On modelling contact and friction—Calculation of tyre response on uneven roads. In: *Proceedings of 2nd Colloquium on Tyre Models for Vehicle Analysis*, Eds. F. Böhm and H.P. Willumeit, *Vehicle System Dynamics*, 27(Suppl.), 1996.
7. Oertel, Ch. and Fandre, A., Ride comfort simulations and steps towards life time calculations: RMOD-K and ADAMS. In: *Proceedings of the International ADAMS User Conference*, Berlin, Germany, 1999.
8. Radt, H.S. and Milliken, W.F., Non-dimensionalizing tyre data for vehicle simulation. *Road Vehicle Handling*, Institute of Mechanical Engineers (C133/83), 1983.
9. Bakker, E., Nyborg, L., and Pacejka, H.B., Tire modeling for use in vehicle dynamics studies. *SAE*, 1987, Paper No. 870421.
10. Pacejka, H.B. and Bakker, E., The Magic Formula tyre model. In: *Proceedings of the 1st Colloquium on Tyre Models for Vehicle Analysis*, Delft, the Netherlands, 1991, Ed. H.B. Pacejka, *Vehicle System Dynamics*, 21(Suppl.), 1993.
11. Bayle, P., Forissier, J.F., and Lafon, S., A new tyre model for vehicle dynamics simulation. In: *Proceedings of the Automotive Technology International*, Aachen, Germany, 1993, pp. 193–198.
12. Pacejka, H.B., The tyre as a vehicle component. In: *Proceedings of XXVI FISITA Congress*, Ed. M. Apetaur, Prague, Czech Republic, 1996.
13. van Oosten, J.J.M. and Bakker, E., Determination of magic tyre model parameters. In: *Proceedings of the 1st Colloquium on Tyre Models for Vehicle Analysis*, Ed. H.B. Pacejka, Delft, the Netherlands, 1991, *Vehicle System Dynamics*, 21(Suppl.), 1993.
14. de Vries, E.J.H. and Pacejka, H.B., Motorcycle tyre measurements and models. In: *Proceedings of the 15th IAVSD Symposium on the Dynamics of Vehicles on Roads and Tracks*, Ed. L. Palkovics, Budapest, Hungary, 1997, *Vehicle System Dynamics* 28(Suppl.), 1998.
15. von Schlippe, B. and Dietrich, R., *Das Flattern eines bepneuten Rades*. Bericht 140 der Lilienthal Gesellschaft, 1941: NACA TM 1365, 1954.
16. Higuchi, A., Transient response of tyres at large wheel slip and camber. Dissertation, TU Delft, Delft, the Netherlands, 1997.
17. van der Jagt, P., Pacejka, H.B., and Savkoor, A.R., Influence of tyre and suspension dynamics on the braking performance of an anti-lock system on uneven roads. In: *Proceedings of the EAEC Conference*, Strassbourg, Germany, C382/047 IMechE, 1989.
18. Pacejka, H.B. and Besselink, I.J.M., Magic formula tyre model with transient properties. In: *Proceedings of 2nd Colloquium on Tyre Models for Vehicle Analysis*, Eds. F. Böhm and H.P. Willumeit, Berlin, Germany, 1997, *Vehicle System Dynamics*, 27(Suppl.), 1996.
19. Zegelaar, P.W.A., The dynamic response of tyres to brake torque variations and road unevennesses. Dissertation, TU Delft, Delft, the Netherlands, 1998.
20. Maurice, J.P., Short wavelength and dynamic tyre behaviour under lateral and combined slip conditions. Dissertation, TU Delft, Delft, the Netherlands, 2000.
21. Schmeitz, A.J.C. and Pacejka, H.B., A semi-empirical, three-dimensional, tyre model for rolling over arbitrary road unevennesses. In: *Proceedings of 18th IAVSD Symposium on the Dynamics of Vehicles on Roads and Tracks*, Ed. M. Abe, Kanagawa, Japan, 2003, *Vehicle System Dynamics*.
22. Lugner, P., Pacejka, H.B., and Plöchl, M., Recent advances in tyre models and testing procedures. State of the art paper of *19th IAVSD Symposium on the Dynamics of Vehicles on Roads and Tracks*, Milano, Italy, 2005, *Vehicle System Dynamics*, 43.
23. Bandel, P. and Monguzzi, C., Simulation model of the dynamic behavior of a tire running over an obstacle. *Tire Science and Technology, TSTCA*, 16, 2, 1988.

24. Schmeitz, A.J.C., A semi-empirical three-dimensional model of the pneumatic tyre rolling over arbitrarily uneven road surfaces. Dissertation, TU Delft, Delft, the Netherlands, 2004.
25. van Eldik Thieme, H.C.A., Experimental and theoretical research on mass-spring systems. In: *Proceedings of the FISITA Congress*, The Hague, the Netherlands, 1960.
26. Bröder, K., Haardt, H., and Paul, U., Reifenprüfstand mit innerer und äusserer Fahrbahn. *ATZ*, 75(2), 1973.
27. Krempel, G., Untersuchungen an Kraftfahrzeugreifen. *ATZ*, 69(1), 8, 1967. (cf. also dissertation Karlsruhe University, 1965.)

ADDITIONAL RECOMMENDED LITERATURE (cf. [1] FOR MORE)

28. Bernard, J.E., Segel, L., and Wild, R.E., Tire shear force generation during combined steering and braking maneuvers. *SAE*, 1977, Paper 770852.
29. Besselink, I.J.M., Shimmy of aircraft main landing gears. Dissertation, TU Delft, Delft, the Netherlands, 2000.
30. Böhm, F., Der Rollvorgang des automobil-rades. *ZAMM*, 43, T56–T60, 1963.
31. Clark, S.K., A brief history of tire rolling resistance. In: *Proceedings of the 1st Rubber Division Symposia*, Chicago, IL, 1982.
32. Davis, D.C., A radial-spring terrain-enveloping tire model. *Vehicle System Dynamics*, 3, 55–69, 1974.
33. Dijks, A., A multifactor examination of wet skid resistance of car tires. *SAE*, 1974, Paper 741106.
34. Fiala, E., Seitenkräfte am rollenden Luftreifen. *VDI Zeitschrift*, 96, 973, 1954.
35. Guan, D.H., Shang, J., and Yam, L.H., Modelling of tire cornering properties with experimental modal parameters. *SAE*, 1999, Paper 1999-01-0784.
36. Guo, K.H., The effect of longitudinal force and vertical load distribution on tire slip properties. *SAE*, 1994, Paper 945087.
37. Guo, K.H. and Liu, Q., Modelling and simulation of non-steady state cornering properties and identification of structure parameters of tyres. In: *Proceedings of the 2nd Colloquium on Tyre Models for Vehicle Analysis*, Berlin, Germany, 1997, Eds. F. Böhm and H.P. Willumeit, *Vehicle System Dynamics*, 27(Suppl.), 1996.
38. Jianmin, G., Gall, R., and Zuomin, W., Dynamic damping and stiffness characteristics of the rolling tire. *Tire Science and Technology*, 29, 4, 2001.
39. Mastinu, G., A semi-analytical tyre model for steady and transient state simulations. In: *Proceedings of the 2nd Colloquium on Tyre Models for Vehicle Analysis*, Berlin, Germany, 1997, Eds. F. Böhm and H.P. Willumeit, *Vehicle System Dynamics*, 27(Suppl.), 1996.
40. Mousseau, C.W. and Clark, S.K., An analytical and experimental study of a tire rolling over a stepped obstacle at low velocity. *Tire Science and Technology, TSTCA*, 16, 2, 1994.
41. Pacejka, H.B., Analysis of the wheel shimmy phenomenon. In: *Proceedings of the Automobile Division of I.Mech.E.*, 180, Part 2A, Institution of Mechanical Engineers, 1965–1966.
42. Reimpell, J. and Sponagel, P., *Fahrwerktechnik: Reifen und Räder*. Vogel Buchverlag, Würzburg, Germany, 1986.
43. Sakai, H., Study on cornering properties of tire and vehicle. *Tire Science and Technology, TCTCA*, 18(3), 1990.
44. Sharp, R.S. and El-Nashar, M.A., A generally applicable digital computer based mathematical model for the generation of shear forces by pneumatic tyres. *Vehicle System Dynamics*, 15, 187–209, 1986.
45. Smiley, R.F., Correlation, evaluation and extension of linearized theories for tire motion and wheel shimmy. NACA (NASA) Technical Note 4110, 1958.
46. Takahashi, T. and Hoshino, M., The tyre cornering model on uneven roads for vehicle dynamics studies. In: *Proceedings of the AVEC'96, International Symposium on Advanced Vehicle Control*, Ed. H. Wallentowitz, Aachen, Germany, 1996.

18 Pneumatic Tire Models
The Detailed Mechanical Approach

Michael Gipser

CONTENTS

18.1 Introduction 625
18.2 Tire Finite Element Analysis 628
18.3 Mechanical Tire Models of Reduced Complexity, Using Dedicated Software 630
18.4 Interfacing between Suspension and Tire Model 631
18.5 Frequently Used Mechanical Tire Models 633
18.6 FTire Model Family 635
18.6.1 Flexible Ring Tire Model 635
18.6.1.1 Structure Model 636
18.6.1.2 Tread Model 639
18.6.1.3 Thermal Model and Wear Model 641
18.6.1.4 Model Data and Parameterization 643
18.6.1.5 Numerical Aspects and Implementation 644
18.6.2 FETire: Coarse-Mesh FETire Model 646
18.7 Parameterization Tools 650
References 652

18.1 INTRODUCTION

It is needless to state that tire properties and tire models are playing, since the very beginning of simulation, a most essential role in all kinds of vehicle dynamics investigations. However, there is a growing demand on ever better and more detailed models, which essentially widen the range of applicability. Today's simulation technology increasingly extends the classical fields of steady-state or low-frequency vehicle dynamics into regions of higher frequencies—with respect to both road excitation and dynamic tire response, in time and in frequency domain.

At the same time, with ever-growing detailing and accuracy of suspension models, the tire models as well are expected to be more and more accurate. Simulation today is no longer a matter of principle investigations, as it might have been in the past. Simulation today no longer means to predict only an order of magnitude but rather two, three, or more decimals of the result. At least, this is what most users would expect an expensive software package be able to achieve.

Moreover, this accuracy is desired not only in operating points that are covered by respective measurements but also for conditions that need extrapolation. For sure, many users of a multibody system (MBS) suspension model would like the tire model to accurately predict tire forces in complex but relevant situations. Imagine a vehicle, passing an obstacle in a sharp curve when braking at the same time, with an anti-lock braking system (ABS) controller being active. What if in addition, the tire is damaged by the obstacle and loses pressure? And what if all this happens during heavy rain or if the vehicle tends to roll over? No doubt, all these are relevant questions of vehicle dynamics, which are much more than just of academic interest. And simulation experiments will have to provide answers, sooner or later.

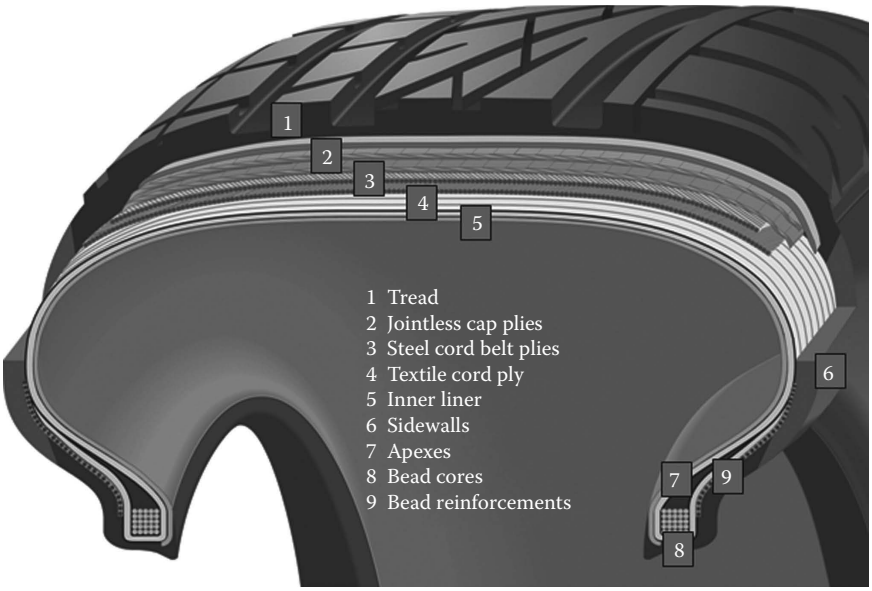


FIGURE 18.1 Components of a passenger car radial tire. (Image courtesy of Continental AG, Hanover, Germany.)

On the other hand, from a modeling point of view, the tire is a tough nut. Its design is a complicated compound of many, very different materials, vulcanized together after an elaborate building process (Figure 18.1). Several physical domains are decisive for its behavior: structural mechanics, thermodynamics, fluid dynamics, chemistry, tribology, heat transfer, and more. From the system dynamics point of view, a tire behaves highly nonlinear. This is not only due to large, up to extreme deflections but also due to nonlinear rubber behavior, unilateral contact forces in the footprint, and friction phenomena during cornering, driving, and braking.

The tread pattern has a large influence on many tire properties. Its geometry ranges from simple (some agricultural, aircraft, and racing tires) to extremely complex (most passenger car tires). The tread pattern of a passenger car tire is a nightmare for those trying to simulate aquaplaning by a combination of mechanical tire models with computational fluid dynamics (CFD) models (Figure 18.2). And it is a nightmare when trying to simulate tread pattern-related noise generation or when trying to predict traction and cornering forces on snow or soft soil.

Putting all this together, by relating the user's demands (wide and ever-growing application range, high frequencies of excitation and response, high absolute accuracy demands, validity and reliability in situations where no respective measurements are available) to the tire's properties (extreme nonlinearity, complicated geometry, complicated material properties, tremendous variety of designs and sizes, comprising tires for bicycles, scooters, motorcycles, passenger cars, SUVs, trucks, trailers, agricultural machines, aircrafts, earthmovers, all-terrain vehicles (ATVs), race cars, and more), it is obvious that simple models will not be sufficient for all this in the long term. Models are required, which orientate themselves as close as possible toward the physics of the tire, but which can be evaluated with bearable computational effort at the same time.

Clearly, it makes a difference whether a tire model will be used in tire industry, to assist in developing and improving the tire itself, or whether it is used by the development department of a vehicle (or a suspension component) manufacturer. The primary interest of the latter will not be to improve the tire but rather to have a description of the tire used or planned for use together with the vehicle under development. As stated earlier, this description has to be as accurate, reliable, and efficient as possible. But it need not necessarily describe the influence of tire design modifications on tire performance.

A tire developer, on the other hand, has (or should have) the best possible knowledge about the tire design and all geometry and material data items that are necessary to generate a detailed finite

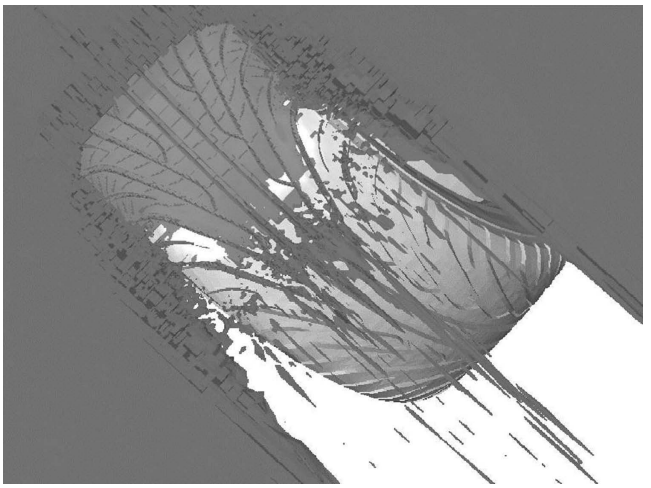


FIGURE 18.2 Hydroplaning simulation using an explicit FEM/FVM code. (Image courtesy of Yokohama Tire, Tokyo, Japan; From Okano, T. and Koishi, M., A computational procedure to predict transient hydroplaning of a tire, *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.)

element (FE) model. FE analysis (FEA) is the most important computer-aided engineering (CAE) tool during tire development. Very much simplified, a tire designer will be ready to spend time and effort to develop FE tire models, because these models are directly related to his primary task and because he has all the available items that are needed.

A vehicle designer, however, typically does not have access to such detailed tire design data. Nor is he or she ready to lose much time in establishing and running an FE model of a product that he is not responsible for. Moreover, the tire model needs to be coupled to his vehicle model. Simultaneously running four FE tire models (or even much more, 12, say, for a typical European tractor–trailer combination), in addition to his vehicle model, would mean unacceptable long computing times. This is the reason why there is a growing demand on tire models, which are

- Based upon a detailed mechanical approach, to be predictive in all relevant situations
- Far less complex than FE models, to be fast enough for full vehicle simulations with standard vehicle dynamics maneuvers in time domain

The biggest challenge for all such models is the supply of model data, subsequently called parameterization. Many of the parameters of a mechanical tire model "below FE" cannot be easily determined or measured directly, even if they have a well-defined physical meaning. So, parameterization tools play a crucial role for mechanical models (as they do of course, even more, for phenomenological models like MF-Tyre and MF-Swift™, which are based upon the famous Magic Formula [1]).

There are two approaches of parameterization. The first uses measurements of

- Static behavior (radial, longitudinal, lateral stiffness, etc.) of a standing tire
- Steady-state behavior (handling characteristics) of a rolling tire
- Modal properties (resonance frequencies and related mode shapes) of an unloaded or loaded tire
- Dynamic response to transient excitation (cleat tests on different types of cleats, at different rolling speeds and vertical loads)

in order to identify the model data. This identification process will be performed or at least assisted by some parameter identification software.

The second approach requires a detailed FE model, either to compute the data of the "coarser" model by applying some static or modal condensation technique or to generate "virtual" measurements of static, steady-state, modal, and transient type.

For both approaches, time and cost effort is considerable. Either sophisticated measuring rigs or detailed FE data are required. Typically, the tire manufacturer seems to be the only place where both are available. Therefore, a suspension designer working for a vehicle manufacturer will try to get a ready-to-run data file from there. However, in turn, the tire model then needs to be accepted and available at the tire manufacturer. This strive for interchange of model data between tire and vehicle manufacturers is one of the reasons why there are only very few tire models actually being used in daily work, as opposed from so many more being published.

18.2 TIRE FINITE ELEMENT ANALYSIS

A survey among leading tire manufacturers revealed that all are intensively applying FEA during tire development. Most companies said the share of FEA among all kinds of CAE computations is more than 80%; some said it is even more than 90%. One manufacturer mentioned their development delays had been cut by 50% in the last 8 years, through applying FEA, with models similar to the one shown in Figure 18.3.

Moreover, nearly all companies being questioned are developing, maintaining, and using proprietary in-house FE software. Apparently, the importance of tire FEA justifies such considerable effort. It seems this is still the easiest way to map the company's tire know-how onto simulation models. Commercially available FE software in use was said to be the ABAQUS™, LS-DYNA™, and PAM™ software family. Few mention NASTRAN™ and Marc™. Clearly, emphasis is put on highly nonlinear models, including large displacement, contact, and friction treatment.

The list of FEA applications being mentioned is long and reflects the complexity of the tire development process. On top of the list are routine calculations like static footprint analyses, comprising shape prediction and computation of ground pressure distribution. The same is performed also during steady-state rolling (Figure 18.4) and, sometimes under certain transient conditions, to predict dynamic wheel loads. Other routine calculations (in order of importance)

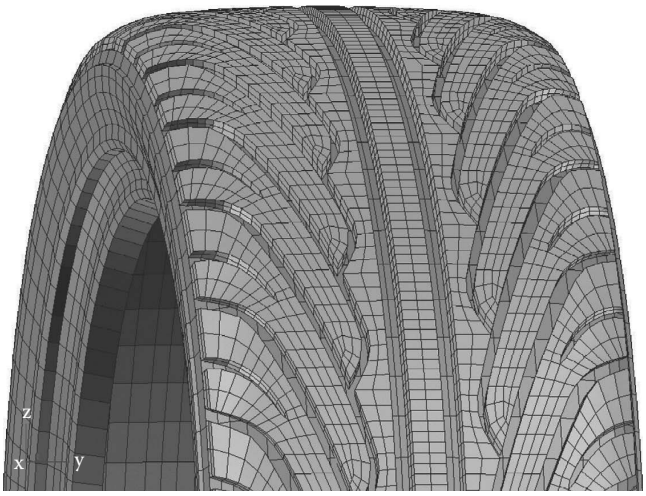


FIGURE 18.3 FE model with medium-scale mesh size. (Image courtesy of Yokohama Tire, Tokyo, Japan; From Okano, T. and Koishi, M., A computational procedure to predict transient hydroplaning of a tire, *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.)

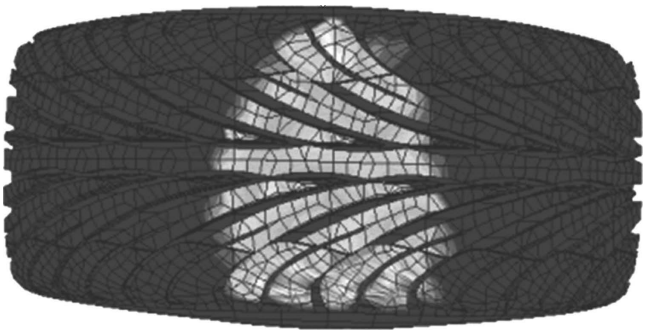


FIGURE 18.4 Pressure distribution at sideslip simulation with in-house FE model. (Image courtesy of Toyo Tires, Osaka.)

are stress and strength in carcass and belt, modal analysis, and prediction of steady-state traction and cornering forces.

More advanced FEA applications, being mentioned by few companies only, are

- Rolling resistance prediction and optimization
- Tire enveloping characteristics on short-waved obstacles
- Tire inflating process and tire–rim fitting including prediction of ease of mounting and dismounting
- Traction on water, snow, and compactable soil
- Estimation of friction energy dissipation, thermal maps, and chemical evolutions
- Tread wear simulation
- Impact and misuse analysis
- Coupled acoustic–structural analysis including excitation process of tread block contact on different ground textures
- Tire building process, including polymer synthesis

Clearly, some of these applications are nonstandard, or nothing more but fundamental research. Several manufacturers had mentioned that making such applications more reliable and cost-efficient will be an important activity in the near future. Other directions of development were said to be

- The integration of isolated FEA runs into an integrated standard procedure to automatically characterize all performance criteria and a further integration of this procedure into the CA process chain
- The improvement of material models for rubber compounds, fibers, and foams, taking simultaneously into account the nonlinear dependency on strain level, excitation amplitude, frequency, load history, and temperature
- Geometrical refinements, going hand in hand with a further increase in computing power
- The improvement of friction models, by trying to better understand thermo-viscoelasticity and taking into account nonstationary friction laws
- The development or improvement of ground models for mud, snow, and sand
- Crack initiation and crack arresting effects by damage simulation
- Simulation of the tire mixing/forming/building process with its consequences on material properties

[2–47] is an incomplete collection of papers dealing with modeling techniques and applications of FEA for tires. More application reports can be found in the journal [48].

18.3 MECHANICAL TIRE MODELS OF REDUCED COMPLEXITY, USING DEDICATED SOFTWARE

In contrast to "true" FE models, which are using a commercial or in-house FEA package, there are a couple of mechanical models under development using software written and optimized for the sole purpose of dynamic tire simulation [49–74]. Models of this kind show reduced complexity, compared to full FE models. Expressed somewhat simplified, the model structure in multipurpose FEA is described merely by the input files, whereas it is "hard-coded" in dedicated software. Such a program might better take full advantage of some tire-specific mechanical and numerical properties. Moreover, it can easily be exchanged across companies, without having available the same FE package at both places, and it can be coupled more easily to MBS suspension models. Using a true FE model would require a complicated co-simulation of FE and MBS solver.

As already stated, there is an ever-growing demand on multipurpose tire model solutions, which are simultaneously applicable in many important fields of vehicle dynamics. Reasons are quite obvious:

- Developing, or purchasing, one model rather than several saves time and money
- Parameterization is to be done only once for each tire size
- Inexperienced users don't need to select "the right" tire model separately for each application
- The model can be used in situations that are influenced by several aspects of vehicle dynamics

Typically, such optimized mechanical tire models use much less degrees of freedom (DOFs) than an FE-based model would. This choice of the number and kind of DOFs has to follow a compromise. The range of applicability, sometimes expressed (in an oversimplified manner) by a maximum frequency value, is to be compared with the implementation effort, the computing time, the numerical stability even in extreme situations, the accuracy, and the effort for parameterization.

As far as range of applicability is concerned, such tire models today are used for

- Handling studies on even and uneven surface
- Ride comfort
- Durability (by generating time-domain load histories)
- Tire misuse (running over deep potholes, say, at high speed)
- Study of tire dynamics interaction with suspension and brake control systems
- Noise, vibration, and harshness (NVH) studies, sometimes using linearized model variants

The tremendously growing computing power of ever cheaper personal computers (PCs) and workstations surely will result in the development and usage of more and more complex simulation models, further widening this range of applications. Let just one number illustrate this power: A standard PC (for about 500 US$) with quad-core central processing unit (CPU) can compute, in every millisecond, far more than one million arithmetic "basic operations" of the form $a = b + c \times d$ with 15-digit precision. In practice, however, there is a risk that non-appropriate numerical algorithms waste a large share of this power.

When specifying the tire model's DOFs, it seems to be quite natural to seek a reasonable balance between tire and suspension model complexity. This complexity can be expressed both in terms of computing time and in terms of model "granularity." In most cases, it will not make much sense to using a zero-DOF point contact tire model together with a highly detailed suspension model, which today typically reaches frequency ranges of 100 Hz and more. Nor is it reasonable to load a complicated FE tire model with only a simple two-mass oscillator.

One of the tire's structural stiffness properties that have the highest impact on the overall tire dynamics is the steel belt's longitudinal stiffness. This stiffness, for example, essentially influences the tire's enveloping characteristics when passing over short obstacles. On the other hand, the stiffness, being quantified by Young's modulus of the belt chord threads, together with the distributed belt mass, is one cause of the numerical stiffness of the resulting simulation model. Such a numerically stiff system can only be integrated either with an explicit solver and an extremely small step size or with an implicit solver. Essentially, the maximum time step h_{max} of an explicit solver is determined by the speed of sound in steel and by the smallest distance of two nodes (i.e., the mesh size Δs):

$$h_{max} \approx \frac{\Delta s}{v_{sound}}$$

The speed of sound of longitudinal waves in thin steel rods is close to 5000 m/s, and a typical mesh size will be less than 20 mm, to resolve even small obstacles. This would result in a maximum step size $h_{max} \approx 4$ μs. Such a value is unrealistically small, if simulations of several seconds or even minutes are to be carried out with the full vehicle.

The use of an implicit solver avoids such a numerical stability restriction. But, on the other hand, this requires to establish and to solve a large linear system of equations. This system is to be updated either in each time step or at least repeatedly after a small number of such steps. The coefficient matrix of the system of equations is closely related to the linearized system matrix of the model, and this matrix in turn is determined by the linearized mass, damping, and stiffness matrix of the mechanical model ([50,56]).

The structure of these matrices is the most important starting point to reduce computing effort. Apart from the geometry of the tread pattern, and some undesirable imperfections, the unloaded tire structure is "nearly" axisymmetric. This axisymmetry can help to considerably reduce computing time for linearization. Even more important, however, is the sparsity pattern of the system matrix (cf. Section 6.1.5). If this sparsity pattern is carefully taken into account in the solution algorithm, computing time can be reduced to only growing linearly with the number of DOFs rather than being a quadratic or even cubic function.

To gain any advantage of these special features, it turns out that tire models will have to perform the integration of their inner state variables by themselves. The numerical integrator becomes an inseparable part of the model software. This approach is followed by most of the commercially available, mechanically based tire models, like FTire™, CDTire™, and RMOD-K™ [71,75].

Being advantageous on the one hand and helping to efficiently solve the model equations, on the other hand, this local integration might turn out to be problematic for the MBS integrator. Most MBS packages are using, by default, complicated, step-size controlling integrators. The accuracy estimation, and thus the step-size control algorithm, might be very sensitively affected by such kind of hidden state variables. As a consequence, the interfacing between MBS packages and tire models plays a crucial role in tire simulation model development. This holds for physical aspects, for numerical aspects, as well as for implementation and coding aspects.

18.4 INTERFACING BETWEEN SUSPENSION AND TIRE MODEL

From the point of view of an MBS program, a tire is nothing but another force element, providing forces and moments acting on the rim, which are depending on all 12 state variables of the rim (Figure 18.5). Of course, the tire's force/moment response is influenced by road attributes, like height and friction coefficient, as well as by the previously mentioned inner states of the tire. In a detailed mechanical tire model, the auxiliary kinematical slip values (longitudinal slip, sideslip angle, and turn slip) are only used as additional output signals. Especially at low speed, tire response

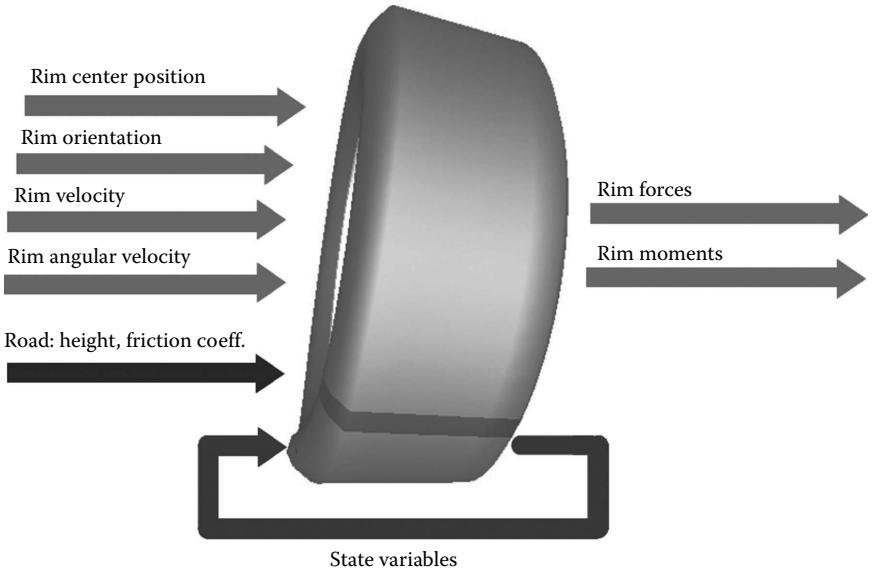


FIGURE 18.5 Tire model interface signals (without ground reaction forces).

cannot be expressed as a function of only these values but rather depends in a more complex way on all rim states at the same time.

If the road surface cannot be considered to be perfectly rigid, tire reaction forces in the contact patch also have to be provided. This might be of importance for heavy truck simulations or aircraft landing gear simulations. Moreover, any snow or soft-soil model to be connected to the tire model will require contact forces as input and react with ground geometry as output. Unfortunately, there is not yet any unified approach for the description of these distributed variables in the tire model program interface. This unification is the precondition to promote further development of soil and tire models independently.

Just like a real tire in the real world, the virtual tire needs to fit to different wheels—the "wheel" in the virtual world being replaced by the calling simulation software. Many users expect that they can use the same tire model together with many different packages capable of simulating motion or dynamics like (in alphabetical order) ABAQUS, FEDEM™, ADAMS™, CarSim/TruckSim/BikeSim™, dSPACE/ASM™, Dymola/Modelica™, MADYMO™, MATLAB®/SIMULINK®/SimMechanics™, MotionSolve™, PAM-CRASH™, RecurDyn™, SAMTECH™, SIMPACK™, veDYNA™, and Virtual. Lab Motion™, as well as on all important operating systems.

There is an activity going on that tries to unify and standardize the program interfaces between MBS packages and tire models—the standard tire interface (STI) of the TYDEX Working Group [76]. This standardized program interface mainly was meant to dock simple tire models with very few and "numerically harmless" states only to a wide variety of MBS packages. This is why the STI does not fit very well to models with many DOFs, being of numerical stiff nature. It is desirable to redesign this interface, using a time-discrete approach, as well as to modernize data formats and coding style.

Despite this necessary revision, such a standardized interface is very useful. It encourages users to compare and evaluate different tire models together with their respective vehicle model. A similar unification is needed for the program interface between tire model and road evaluation routine. The TYDEX Working Group defined a proposal for this task as well ([76]). As opposed to the tire/vehicle interface, this proposal never reached the necessary acceptance. A revision should include time-dependent moving surfaces (like test drums and test rigs with hydraulic actuators) and allow putting out reaction contact forces, to enable soft-soil modeling.

18.5 FREQUENTLY USED MECHANICAL TIRE MODELS

Clearly, it is difficult to gain a comprehensive overview on all tire models in use today. However, only a few are frequently mentioned in application reports and are commercially available. Mechanical models "below" FEs can be roughly categorized into 2D or 3D spokes models, 2D or 3D rigid belt models, and 2D or 3D flexible belt models. The so-called modal models follow a completely different approach.

Spokes models [52,66] can serve as a simple approach to resolve even short-waved road irregularities, if the excitation of structural tire vibrations is not of interest. The application range of this kind of models will be very much limited by this restriction. Spokes are nothing but unilateral spring–damper elements, connected at one end to the rim and occasionally being in sticking or sliding contact to the road at the other end. These spokes replace the stiffness of the tire structure and that of the tread at the same time. The spokes might be placed in the rim midplane only, to simulate pure in-plane events, or might extend over the tread's width as well.

In more realistic models, some kind of inertia is placed between the structure's elasticity and the tread's elasticity. The simplest models of this kind are the rigid belt models. Here, the heaviest vibrating part of the tire structure, the steel belt, is concentrated into one rigid body with three or six DOFs (2D or 3D case, respectively). The spokes are no longer connected to the rim but to this rigid body instead—looking more like a brush than like spokes now. Of course, only the rigid-body modes of the tire structure can be seen as vibrations of such a model. After all, the respective eigenfrequencies can reach the order of magnitude of 100 Hz, depending on the kind of tire.

Models of this kind are

- The author's brush and ring tire (BRIT model, [53,49,63], development abandoned)
- The low-end member of the RMOD-K 6-tire model family ([61,64,71], formerly commercially available)
- The RMOD-K 7 RB member of the RMOD-K 7-tire model family (successor of RMOD-K 6, commercially available)
- RTire (rigid-ring tire model, a "low-end member" of the FTire model family, [77], development abandoned)
- The member CDT20 of the CDTire model family (comfort and durability tire model, [72], commercially available). All model family members of RMOD-K 6 and CDTire have common roots, going back to Böhm [38]

The most widely used rigid belt model, however, seems to be the commercially available

- MF-Swift (short wavelength intermediate frequency tire model, [1])

In the strong sense, MF-Swift is not exactly a model of the type described earlier. In fact, it does not replace the tread by a brush of contact elements. Rather, the horizontal contact forces are not computed on the basis of brush displacements but by a Magic Formula approach. The reaction to short-waved road irregularities is improved by applying a special nonlinear road filter. A very similar approach is implemented in the

- Rigid-ring tire model of Mancosu et al. [70]

If in-plane or out-of-plane flexible belt vibrations are taken into account, both the frequency range and the accuracy can be extended tremendously. All models of this kind replace the rigid-ring belt model by a lumped-mass system. Typically, some 50–2000-point masses make up such a system; there is a smooth transition to coarse-mesh FE models. The masses are connected between each other and to the rim by a system of discrete, mostly nonlinear springs (or generalized elasticities),

and by certain damping elements. Several models also take into account internal forces along the inner liner surface normal, due to inflation pressure. As an example, the author's

- FTire (flexible belt tire model, [75,78–84], commercially available)
- FETire (coarse-mesh FE tire model, commercially available)

are discussed in the succeeding text. The latter is based upon

- DNS-Tire (dynamical nonlinear spatial tire model, [49,50,55])

Other models of this kind are

- The high-end member of the RMOD-K 6-tire model family
- The RMOD-K 7 FB member of the RMOD-K 7-tire model family
- The members CDT30 and CDT40 of the CDTire model family
- The model of Mousseau et al. [68], implemented as subsystem in ABAQUS

The RMOD-K 7 FB (flexible belt) model (Figure 18.6) is quite similar to FETire, both in model structure and in the level of detail.

Quite different from the earlier listed "lumped parameters" models are modal models, an example being the

- Flexible body tire model of Mancosu et al. [68]

This model describes the tire structure as a general flexible body within ADAMS. The actual shape of such a flexible body is defined by a linear combination of the first low-frequency mode shapes of

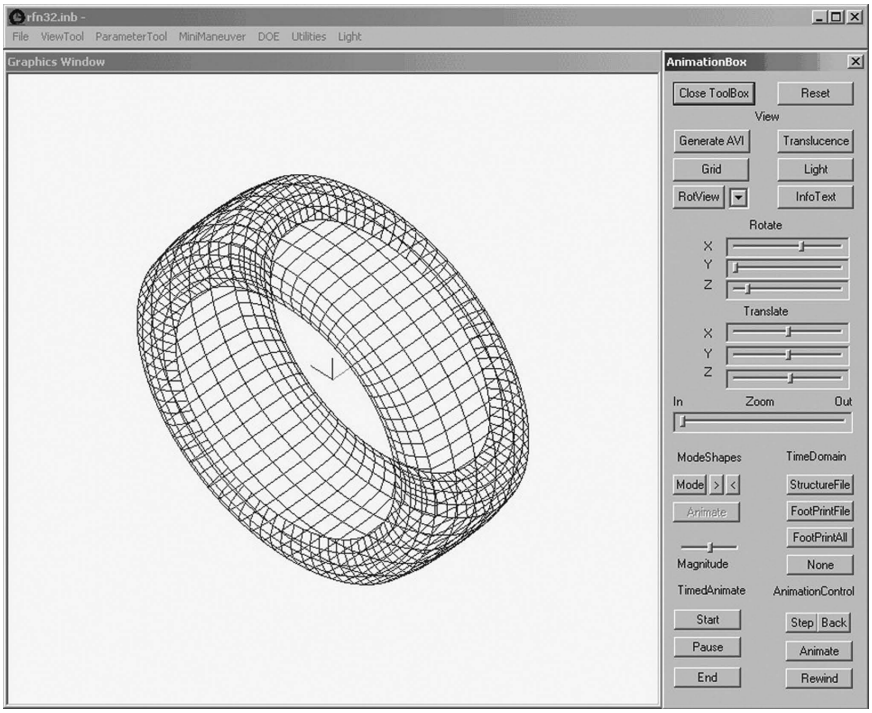


FIGURE 18.6 RMOD-K™ 7 pre- and postprocessor. (Image courtesy of IAT Dynamics GmbH, Berlin.)

the unloaded tire. The coefficients of this linear combination are the DOFs of the structure model. The road contact is established in a similar way than with the lumped parameters models. Such a flexible model requires the availability of a respective full FE model, to provide the modal data. An advantage of the approach is the exact reproducibility of arbitrary eigenfrequencies and related mode shapes. A disadvantage, however, is the intrinsic linearity of the structural stiffness, which will be violated for large displacements. As one consequence, the extremely small circumferential extensibility of the steel belt cannot be guaranteed when rolling over high obstacles.

18.6 FTire MODEL FAMILY

Being a typical representative of detailed mechanical tire models "below" true finite element method (FEM) models, the author's model family FTire (Flexible Ring Tire Model) will be discussed in greater detail here. FTire development started in 1998, using certain ideas and numerical concepts of the author's "coarse-mesh" FE model DNS-Tire [49,50,55] and the spatial nonlinear "rigid-ring/brush" model BRIT [53,59,63]. This first version of FTire has been essentially improved since then. It is available as advanced tire model of ADAMS, SIMPACK, and MotionSolve and is as well implemented into several other commercial and noncommercial simulation programs and as a blockset in MATLAB/SIMULINK. FTire is completed by a redesign of DNS-Tire (called FETire) and a compatible implementation of the 2002 version of the Magic Formula model, called HTire. FTire, FETire, and HTire together set up the FTire model family. FTire and FETire are presented in greater detail in the succeeding text.

18.6.1 Flexible Ring Tire Model

FTire is the most important member of the FTire model family. Like most other mechanical tire models, the kernel of FTire consists of two separate parts. The first part describes the tire's structural stiffness, damping, and inertia properties (the "structure model," for short). The second one is the tread/road contact model, comprising road evaluation and computation of contact pressure distribution and distributed friction forces (subsequently called "tread model").

The underlying idea of the structure model is to introduce only as few DOFs as necessary, to represent all tire properties that are relevant for the application. Main objectives during development had been

- Fully nonlinear 3D model, working in time domain
- Valid up to 150 Hz both in-plane and out-of-plane
- Both longitudinal and lateral wavelengths ≤ 2 cm, including sharp-edged obstacles
- Parameterization as easy and flexible as possible
- Usage of "easy-to-measure" static, steady-state, and modal properties in parameterization
- Computing time $\leq$5 to 20 times real time, and even real-time capable in an optimized version
- Valid for ride and handling, both in steady-state and dynamic conditions
- Simple and flexible interfacing to vehicle and road models

Potential applications of FTire comprise

- Tire models for passenger cars, motor cycles, race cars, trucks, trailers, ATVs, aircrafts, earth-moving machines, etc.
- All kinds of vibrations up to about 150 Hz, induced by tire/road contact
- Vibration excitation through tire imperfections
- Generation of load histories for durability simulations
- Traction and handling on extremely uneven roads

- Rolling resistance studies on even and uneven road
- Tire misuse in different conditions
- Steering torque amount when parking
- Assessment of highly dynamic suspension control systems
- Moving ground, all kinds of test-bench simulations
- Tire temperature and its influence on road friction
- Tread wear
- Pressure loss and tire puncture of standard and run-on-flat tires
- Sinkage and traction on soft soil
- Influence of elastic and viscoelastic rim deformation under large vertical and horizontal loads
- Air cavity vibration

Most model extensions, like thermal and wear model, tire misuse, tire imperfections, and flexible rim model, have been stimulated by users in vehicle and tire industry. Of course, there are still many effects that are of relevance in suspension design and optimization, but cannot be analyzed using FTire, like noise generation and hydroplaning.

18.6.1.1 Structure Model

The FTire structure model consists of several lumped-mass nodes (typically some 60–360), replacing the tire's steel cord. These nodes are connected to the rim and to each other by several nonlinear, inflation pressure-dependent stiffness, damping, and friction elements. Moreover, the nodes are subject to inflation pressure forces in radial direction and to the forces of the tread model. The part of the tire structure that is associated to one such belt node is called belt segment. Each belt segment has $4 + x$ DOFs (Figure 18.7):

- Longitudinal, lateral, and vertical displacement
- Rotation angle about the circumferential axis ("belt torsion")
- Bending in lateral direction, perpendicular to the circumferential axis (described by x independent eigensolutions of the bending beam equation)

The elastic coupling between belt nodes and rim is described by nonlinear force elements in radial, circumferential, and transversal direction, comprising springs, dampers, Maxwell elements, and spring-friction series connections ("elasto-plastic" elements). Figure 18.8 shows the radial force element of one single belt node; circumferential and transversal elements are placed analogously. Maxwell elements are used to describe the "tire stiffening" at higher rolling speeds, whereas the dry friction element replaces rubber hysteresis. The effect of this hysteresis on overall tire radial stiffness can clearly be seen in measurements, as well as corresponding FTire simulations (Figure 18.9).

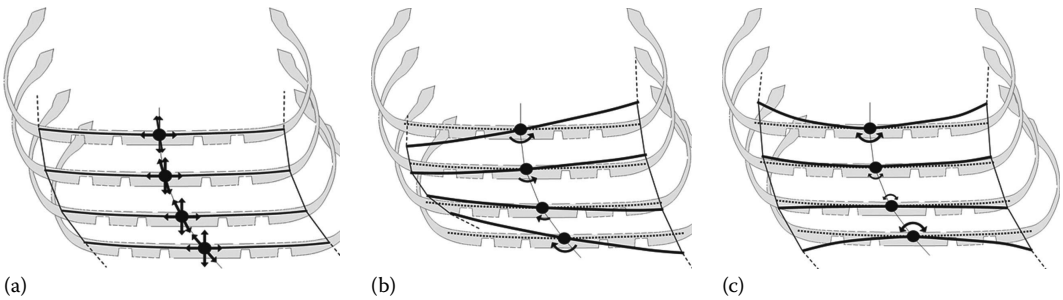


FIGURE 18.7 Belt segments' DOFs (degrees of freedom): (a) translation, (b) torsion, (c) lateral bending.

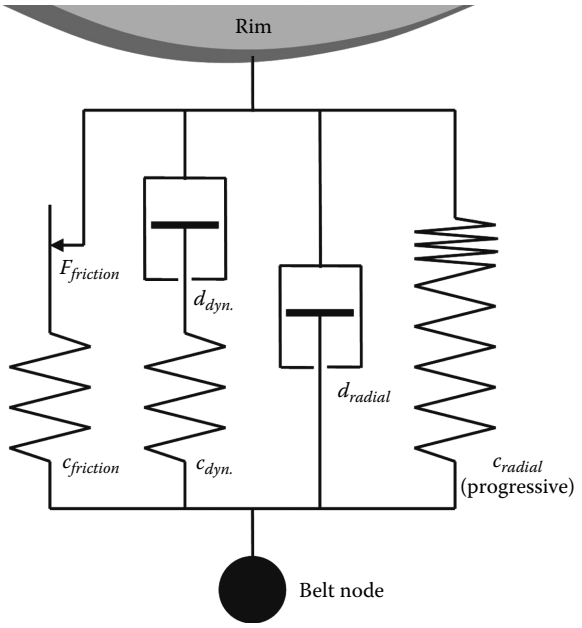


FIGURE 18.8 Radial force elements between each belt node and rim.

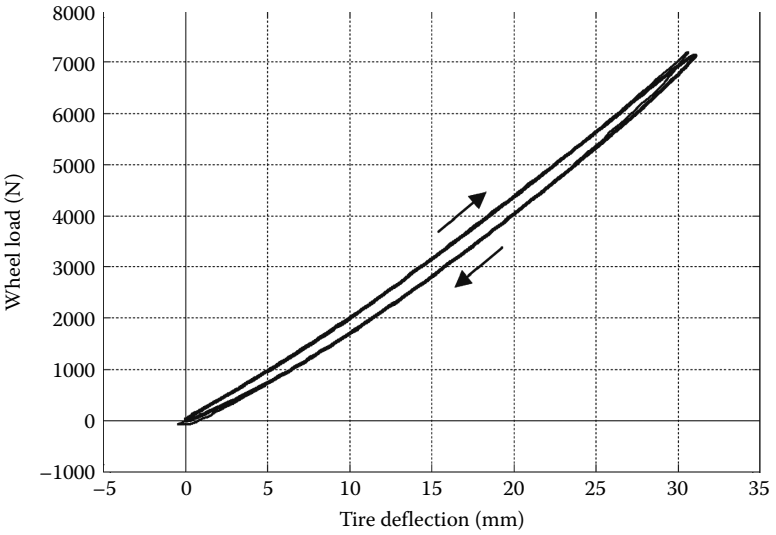


FIGURE 18.9 Measurement (bold line) and FTire simulation (thin line) of radial tire characteristic with hysteresis.

The radial stiffness elements are completed with unilateral stiff springs, located along the radial direction at the rim flanges. When tire deflection exceeds a certain value, these springs will come into contact with the inner side of the belt, thus generating a highly progressive radial stiffness characteristic (Figure 18.10). Depending on camber angle, this rim-to-belt contact can happen on one rim flange only, or on both at the same time. Rim-to-belt contact enables to simulate certain tire misuse situations.

In addition to the translational force elements, a torsional spring $c_{torsion}$ (cf. Figure 18.11) is located between each belt segment and the rim. These torsion springs serve as "elastic foundation"

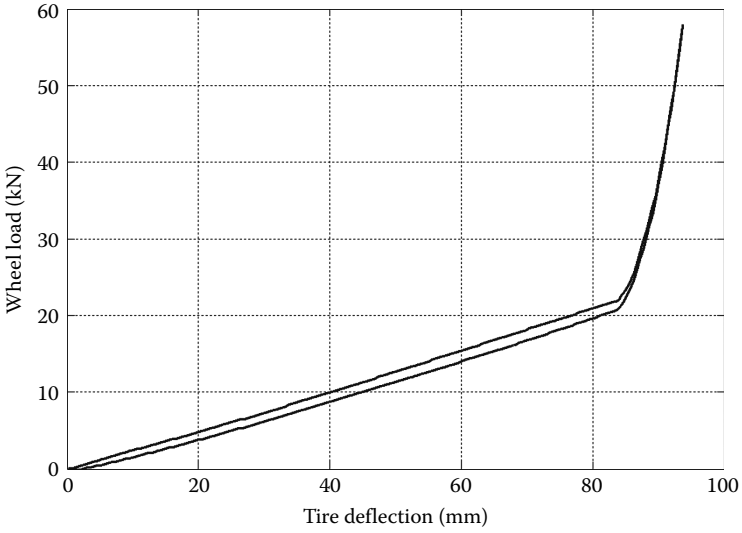


FIGURE 18.10 Progressive radial stiffness due to rim flange contact.

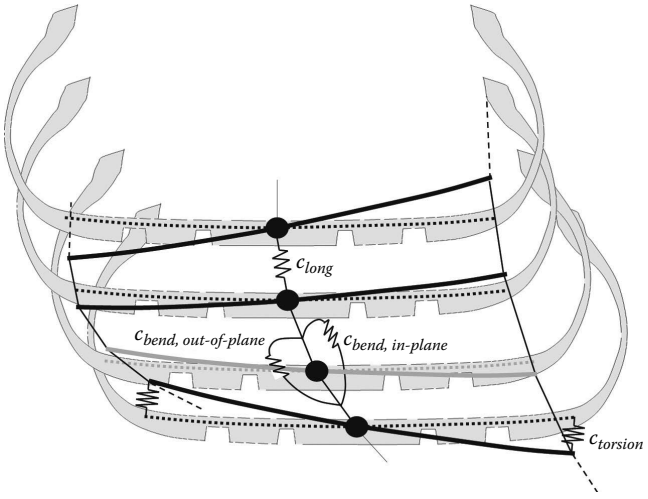


FIGURE 18.11 Belt flexibility stiffnesses.

of the belt segments' torsion DOF. Moreover, a bending stiffness acts against lateral belt bending. Together with the radial stiffness, they describe left and right sidewall stiffness.
The following additional force elements are connecting neighboring belt nodes:

- A very stiff translational spring c_{long} (Figure 18.11) providing the belt extension stiffness
- Another torsional spring, connecting the torsion DOF of two adjacent nodes (this spring is not depicted in Figure 18.11)
- A bending stiffness $c_{bend,in\text{-}plane}$ (Figure 18.11), connecting three nodes in line, acting against belt bending perpendicular to the lateral axis ("in-plane bending stiffness")
- Another bending stiffness $c_{bend,out\text{-}of\text{-}plane}$ (Figure 18.11), acting against belt bending perpendicular to the radial axis ("out-of-plane bending stiffness")

All stiffness values depend on the actual inflation pressure. The pressure is treated as "operating condition" and might be arbitrarily changed during a running simulation. This will not disturb the

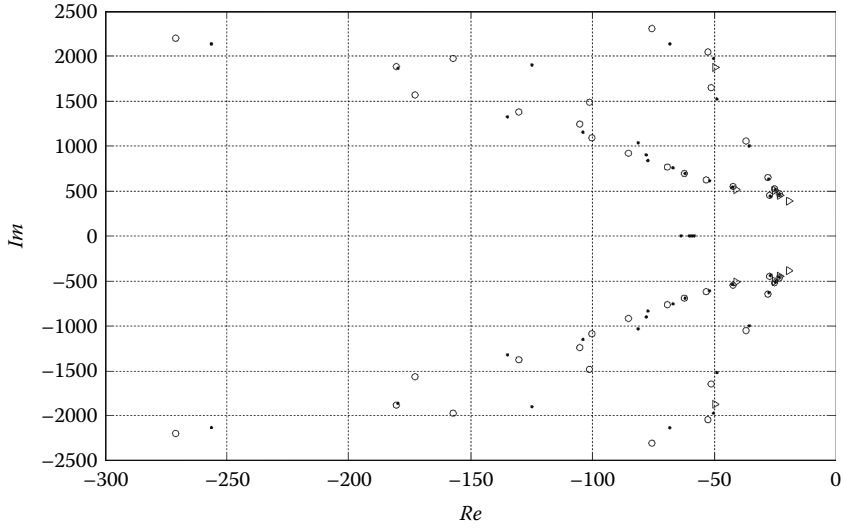


FIGURE 18.12 FTire eigenvalue location of linearized structure model: ▷, unloaded tire; ■, 20 mm deflection; ○, 40 mm deflection; real and imaginary axes are scaled differently.

integration process in any way. Moreover, actual inflation pressure depends on tire temperature, which is the output of the thermal model (cf. Chapter 6.1.3). Inflation pressure does not only affect the stiffness values but generates extra forces, which act on the belt nodes in radial direction. These forces generate a membrane tension, by expansion of the longitudinal springs. This leads to a further "stiffening" of the steel belt, similar to that one caused by the in-plane bending stiffness.

Figure 18.12 shows the first eigenvalues of the linearized structure model, for three different radial deflection values. The linearization technique applied here can be used to compute transfer functions under different operating conditions as well.

Several optional types of imperfections, to better match a real tire, complete the structure model. These imperfections comprise static and dynamic imbalance, radial and tangential nonuniformity, ply steer, conicity, and geometrical runout.

18.6.1.2 Tread Model

Between two neighbored belt segments, there are placed a certain number of massless contact and friction elements, to establish road contact (Figure 18.13). Their number, typically 30–100, is selected by the user, depending on the desired road irregularity resolution. These contact elements constitute the brush-type tread model.

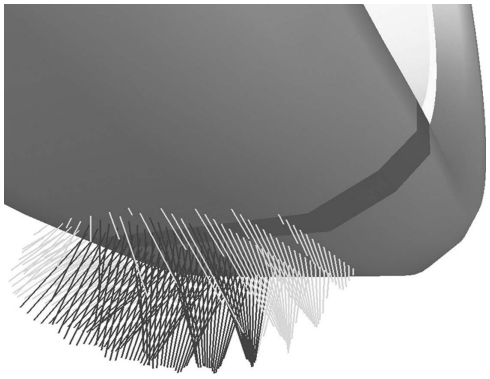


FIGURE 18.13 Distributed contact forces during parking maneuver.

The contact elements are placed in such a way that both the longitudinal and the lateral resolution, defined by the minimum distance of two elements, are optimized with as few as possible elements. The exact positions, in global coordinates, at which the contact points are attached to the distorted belt segments are smoothly interpolated, using the coordinates of the four nearest-by belt nodes, as well as the actual belt's cross-sectional geometry. Vice versa, the forces that are generated by the contact elements will contribute to the external forces of either of these four belt nodes. The actual belt's cross-sectional geometry is defined by all DOFs of the two nearest-by belt segments.

If the tread pattern is given in terms of a black-and-white bitmap file, the contact elements' lengths (i.e., the local tread rubber height) may be set according to this pattern. In Figure 18.14, the computed contact pressure of such a model is shown. Here, 120 segments and 50 contact elements had been used; the contact elements are placed along 25 parallel lines. This resulted in a tread resolution of approximately 0.5 × 12 mm for the large truck tire. Clearly, it would require a much larger number of contact elements to resolve more complicated patterns, like those of typical passenger car tires. However, as demonstrated in the succeeding text, computing time grows less than linearly both with the number of contact elements and the number of belt segments.

During any integration step, for each contact element, a rapid test is performed to check whether the element might have road contact. This test uses the position of the nearest-by belt element. If the test is passed, the exact distance of the contact element to the road surface is computed. Only if the contact element penetrates the road, the contact computation is continued. The next steps then are determining the road surface tangential plane, the element's normal deflection, the resulting normal force, the tangential sliding velocity, and finally the resulting friction forces.

The road tangential plane is computed individually for each contact element, by evaluating the road height in three different locations near the contact element. This is necessary to resolve even sharp-edged obstacles, like cleats and potholes. Road surface may depend on time, like it does for four-post test rigs and rotating drums. In that case, both normal and tangential surface velocities are taken into account as well.

The normal force is a function of deflection and deflection velocity, describing tread rubber compression stiffness and damping. These values, in turn, are determined on the basis of the tread rubber's Young's modulus, the tread pattern's net-to-gross ratio, the local tread depth, and more. Clearly, pressure distribution and thus normal force strongly depend on the tire's cross-sectional geometry, most of all on the belt curvature in unloaded condition as well as on tread depth as function of the lateral coordinate. This is why this geometric data can be prescribed in a very detailed way, by using lookup tables with smooth spline interpolation.

After having computed the normal force, in the next step the vector-valued friction force in the tangential plane is determined. Its amount is given by

$$\left|F_{friction}\right| = \mu(v_{slide}, p_{ground}, T_{tread}) \cdot F_{normal}$$

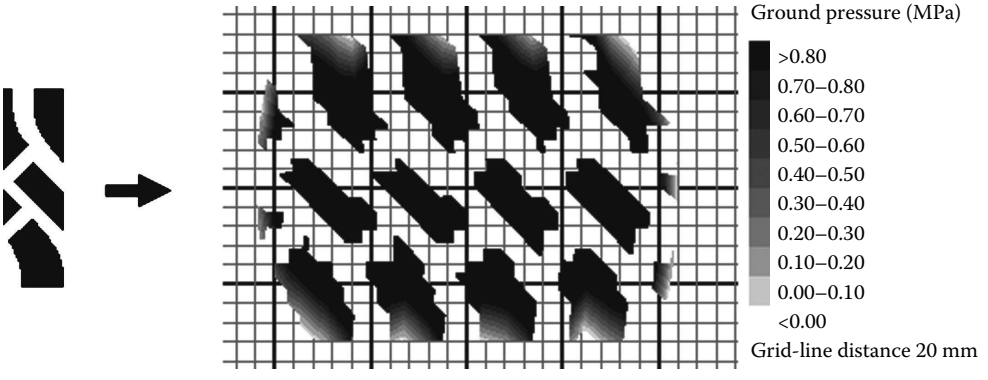


FIGURE 18.14 Tread pattern bitmap and resulting FTire contact pressure distribution.

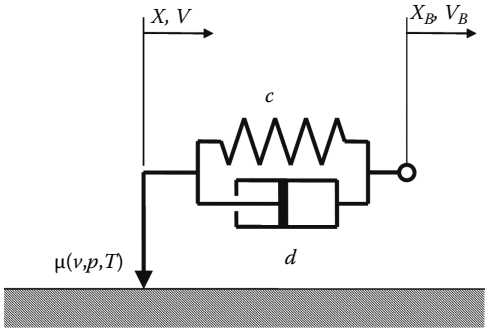


FIGURE 18.15 Contact element tangential displacement model.

and its direction is the negative sliding velocity direction. Sliding velocity is determined by using the vector-valued force equilibrium condition of friction force, elastic shear forces, and shear damping force of the contact element (Figure 18.15), in the road tangential plane.

This force equilibrium condition results in a differential equation for the element's tangential displacement. Depending on the nature of the friction coefficient as function of sliding velocity, this differential equation might have a discontinuous right-hand side or even might be locally unstable. The only way to numerically solve such an equation is using implicit integration. It turns out that this integration requires an additional discrete state variable, which memorizes whether the element, in the previous time step, was sticking to ground or sliding.

18.6.1.3 Thermal Model and Wear Model

The tread model is completed by a thermal model and a by model for tread wear estimation. The thermal model consists of

- The thermodynamic computation of the actual inflation pressure as function of filling gas mass, "cold tire inflation pressure," tire temperature, and actual interior volume.
- A heat generation and transfer model, introducing state variables for the temperature of the tire structure (including filling gas), and the individual temperature of each tread contact element. Heat generation and heat transfer are driven by the power loss distribution due to structural damping and dry friction on the road surface. As already mentioned, tread temperature in turn influences the road friction coefficients.

In the heat generation and transfer submodel, the tire is assumed being separated into three regions, having different thermal properties each (Figure 18.16): the tire structure (including bead, sidewalls, belt, and air volume), the tread without contact patch, and the contact patch. The following assumptions are made for the three different regions.

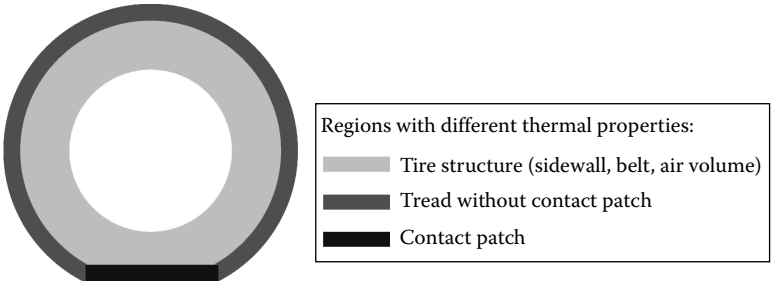


FIGURE 18.16 Regions of the thermal model.

The tire structure is described by one global temperature only. The product of rate of change of this temperature and the tire structure's overall heat capacity is balancing the sum of

- The power loss in all energy dissipating force elements in the belt and sidewall, excluding friction and damping of the tread elements.
- The heat that flows from the tire structure into the two tread regions. This heat transfer is determined by the respective temperature differences, multiplied by an appropriate heat transfer coefficient. This heat transfer coefficient is assumed to be independent on rolling speed.
- The heat that is transferred from the sidewalls into the air flowing around the tire. This transfer is determined by the temperature difference between tire structure and ambient air, multiplied by an appropriate heat transfer coefficient. This heat transfer coefficient is assumed to be strongly and nonlinearly dependent on rolling speed.

A distributed temperature field is assigned to the part of the tread outside the contact patch, providing one temperature value individually for each contact element. The product of rate of change of this temperature and the tread element's heat capacity is balancing the sum of

- The power loss in the tread element due to material damping.
- The share of the heat that flows from the tire structure into the particular tread element, as described earlier.
- The heat being transferred from the tread element to the air flowing around the tire. This heat transfer is determined by the temperature difference between tread element and ambient air, multiplied by an appropriate heat transfer coefficient. Again, this heat transfer coefficient is assumed to be strongly and nonlinearly dependent on rolling speed but has the same value for all elements not in contact to the road.

The contact patch is described in a similar way like the remainder of the tread, with the following exceptions: additional power loss due to road friction is taken into account, and heat is transferred to the road instead of the surrounding air. Heat transfer coefficient between tread and road is assumed to be independent on rolling speed.

As an example, Figure 18.17 shows the resulting temperature distribution in contact patch, after 2 s cornering at 4 kN wheel load, 4° camber angle, and 8° sideslip angle.

The tread wear model uses the friction power (which is the product of friction force and sliding velocity) and a functional relationship between friction power and wear rate. These variables are used to update, individually for each contact element, the actual local tread depth:

$$\dot{d}_{tread} = -f_{wear}\left(\left|F_{friction}\right| \cdot \left|v_{slide}\right|\right)$$

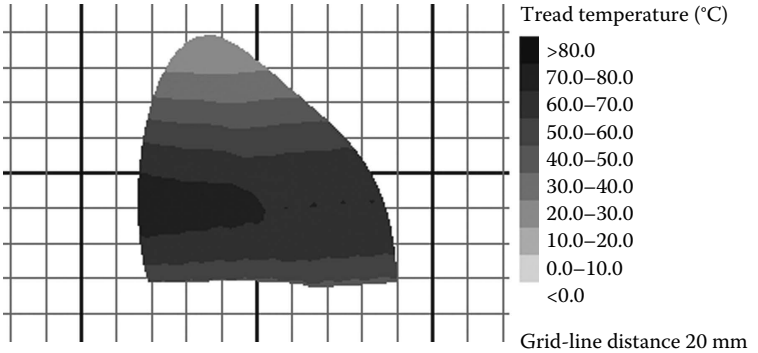


FIGURE 18.17 Temperature distribution in contact patch during cornering at 6° camber angle.

Accompanied by this reduction in tread depth, all tread properties that depend on tread depth will be modified respectively: compression stiffness and damping, shear stiffness and damping, heat capacity, and tread mass distribution. Thus, even imbalance and runout caused by a locked braking will be seen as result of a respective simulation experiment. Clearly, both the thermal and the wear model can be deactivated. Constant temperature and constant tread depth will be used in that case.

18.6.1.4 Model Data and Parameterization

As already discussed in Chapter 1, ease of parameterization is one of the most important objectives of any tire model, and this of course especially holds for FTire. Here, a clear distinction is made between data used in the model equations ("preprocessed data") and data to be supplied by the user ("basic data"). The underlying idea is to define basic data, which are as easy as possible to obtain and which at the same time yield complete information to determine the preprocessed data in a unique way.

This preprocessing requires the solution of several nonlinear systems of equations and will take some seconds of computing time. After successful completion, the resulting preprocessed parameters are appended to the data file. When using this data file the next time, it will be checked whether basic data have been changed in any way. If not, preprocessing phase is skipped, using the appended data instead.

Speaking about effort and cost of measurement-based parameterization, obviously it is hard to say whether certain parameters are "easy to obtain" or not. What is "easy" at one test facility might be difficult or impossible to get at another one. This is why there are different alternatives and combinations for some of the basic data, especially for that of the structure model. Moreover, not all data items are equally relevant for all applications. For example, in many cases, belt out-of-plane bending stiffness does not affect the tire response when rolling over a transversal cleat, and the friction coefficients nearly do not influence the radial tire characteristic.

Besides elementary data like tire and rim size, load index, speed symbol, mass, and rolling circumference, the structure model requires one out of several different combinations of static and modal data of the total tire. During preprocessing, the corresponding internal stiffness, damping, and inertia data are determined such that the resulting FTire model shows exactly this prescribed global behavior.

The following data enumeration only reflects one such possible combination. A comprehensive and up-to-date list of all data items can be found at www.cosin.eu. Tools to assist in parameterization will be discussed in Chapter 7.

A typical FTire parameterization might use

1. Tire size, load index, and speed symbol
2. Rolling circumference
3. Tire mass
4. Rim diameter and rim width
5. Tread width and tread depth as function of lateral tread coordinate
6. Distance of tread grooves to steel belt ("cap base height")
7. Tread pattern net-to-gross ratio (alternatively, tread pattern bitmap file)
8. Mean lateral belt curvature radius of the unloaded tire's cross section (alternatively, detailed spline data of cross-sectional geometry)
9. Young's modulus of tread rubber
10. Tread rubber adhesion friction coefficient for one up to three contact pressure values
11. Tread rubber sliding friction coefficient for two sliding velocities and one up to three contact pressure values
12. Footprint contour and footprint pressure distribution at different vertical loads, camber angles, and inflation pressure values

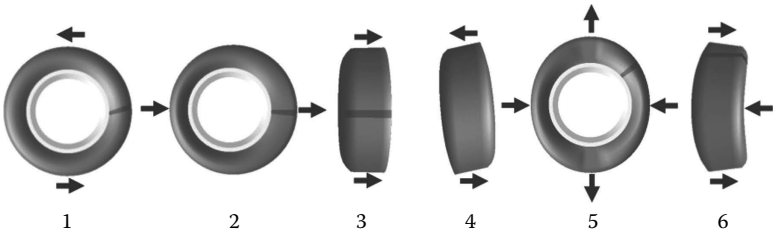


FIGURE 18.18 First unloaded vibration modes for use in FTire parameterization and validation.

13. Natural frequencies and modal damping of vibration modes 1, 2, 4, and 6 (cf. Figure 18.18), for an unloaded, inflated tire with fixed rim, for one or two inflation pressure values
14. Vertical force:
    a. Of standing tire on a flat surface, for two deflection values and one or two inflation pressure values
    b. Of standing tire, on both a transversally and a longitudinally oriented cleat, for one or two inflation pressure values
    c. Of standing tire at large camber angle, on both a flat surface and a transversally oriented cleat, for one or two inflation pressure values
    d. Of rolling tire on flat surface, for two different rolling speeds
15. Horizontal forces of a blocked tire, pulled in longitudinal or lateral direction until it starts sliding

18.6.1.5 Numerical Aspects and Implementation

As discussed in Chapter 4, tire models need to be very versatile with respect to the calling MBS software and the computing environment. For that reason, FTire is implemented in a mixture of Fortran 90 and C, making best use of both the still outstanding "number-crunching" features of modern and standardized Fortran and the system-focused C functionality like multithreading, library plug-in technology, interfacing to OpenGL graphics, data piping, and function pointers.

The interface to the calling MBS solver is organized in terms of a C++−, C−, and Fortran-compatible application programming interface (API) and compiled for all 32 and 64 bit Linux, Windows™, and Mac OS X™ operating systems. Seemingly arbitrarily many instances of the model can be simulated in parallel, in a thread-safe manner and without interfering with each other.

The numerical integration basically is performed in two steps for each time increment. This time increment might be of fixed or variable length. It will be automatically subdivided into smaller steps if rim rotation increment exceeds a certain angle threshold (1°, say). Thus, FTire can be easily linked to MBS software, which uses step-size controlling integrators and yields sufficient accuracy independent on rolling speed. As with all highly dynamic subsystems, the MBS solver's integrator should be configured to limit the communication interval between MBS model and tire model to a reasonable value. In the case of FTire, a value less or equal to 1 ms is recommended. Otherwise, the MBS model cannot accurately follow any high-frequency tire excitation. Just one number might illustrate this: the tire's total contact time to a cleat, ran over at a speed of 200 kph, is not more than 4 ms. Even though FTire accumulates the forces of the internally used smaller steps, the feedback due to the change in rim velocity is decisive as well.

In the first step of each local time increment, the contact processor is called. It will perform all necessary updates of the tread model-related state variables and compute the generalized forces that act between tread rubber and belt segments. The contact processor also calls the road evaluation routine, to determine location—and time-dependent road height and friction value modification. A simple program interface to these evaluation routines was chosen and implemented, to enable the connection of FTire to a wide variety of user-written and application-specific road implementations.

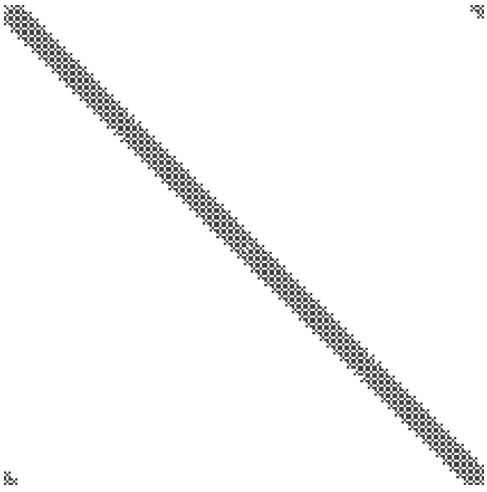


FIGURE 18.19 Sparsity pattern of stiffness matrix.

In the second step, the tire structure model is updated. The size of this system is determined by the number of the belt segments' DOFs. Integration is done by a slightly modified implicit trapezoidal scheme, which in turn is a special case of the widely used Newmark integrator. This integrator requires certain system Jacobian: the matrices of partial derivatives of acceleration variables with respect to position and velocity states. These matrices, being closely related to the linearized structural stiffness and damping matrices, are extremely sparse. Moreover, even though they depend on the actual values of the tire structure states, they can be computed analytically in only short extra computing time. Figure 18.19 shows the nonzero pattern ("sparsity pattern") of these matrices, which are the so-called cyclic band matrices, in the case of 80 belt segments. Only 3.9% of all matrix elements are nonzero. Even after the Cholesky factorization, this value is only about 7.8%. This is important because the number of arithmetic operations to solve the system is closely related to the number and location of nonzero elements of the Cholesky factor.

Figure 18.20 shows the resulting measured CPU time of FTire, in dependency on both the longitudinal contact element distance (which determines the spatial contact patch resolution) and the number of

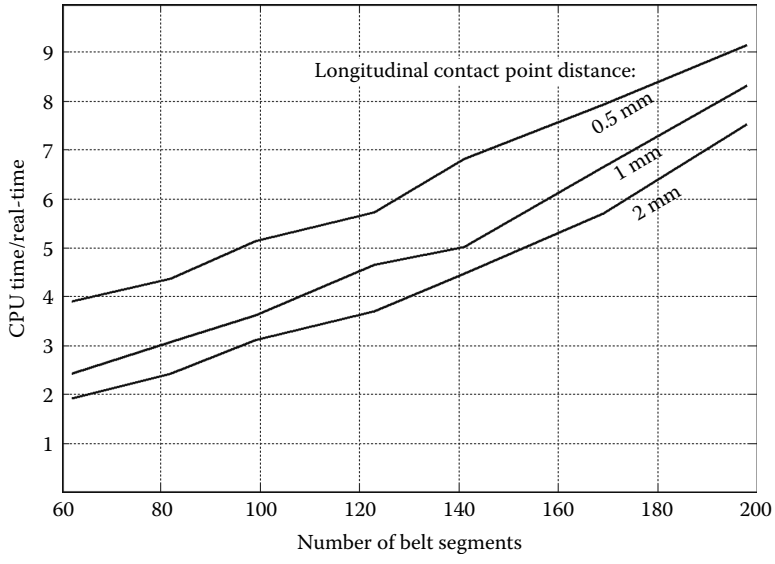


FIGURE 18.20 CPU time of FTire, depending on belt segments and contact patch resolution.

belt segments. Computations had been performed on a standard PC. The diagram shows the real-time factor (RTF), which is the quotient of CPU time to real time. Apparently, computing time depends nearly linearly both on the number of belt segments and on the number of contact elements. For a reasonable numerical tire discretization (60 belt segments and 1 mm contact patch resolution), FTire even can be optimized to reach real-time computation speed and consequently is available in a real-time version.

More about FTire, the complete documentation, required experimental data, and some applications and validation results can be found at [77].

18.6.2 FETire: Coarse-Mesh FETire Model

As mentioned earlier, FTire is completed with a "coarse-mesh" FE model (FETire) and a compatible Magic Formula 2002 implementation (HTire). All the members of this FTire model family, being of very different complexity, use exactly the same program interfaces. At present, FTire is understood as the most important member of the model family, because it seems to be the best compromise between application range, accuracy, and computing time for most vehicle dynamics investigations.

FETire, on the other hand, requiring about 50–100 times more CPU time than FTire, today is mainly used for

- Assistance in the parameterization of FTire
- Theoretical investigations on tire response to "nonstandard" excitations
- Studies about influence of tire design characteristics to handling and ride comfort properties

However, with ever-growing computing power and complexity of suspension models, FETire might be used more widely in the future, even with full vehicle models. RTF is about 200 at present, so 10 s multithreaded simulation of a full vehicle, completely equipped with FETire, would take about 33 min CPU time. This might be a tolerable effort for occasional or special investigations.

Similarly as FTire, FETire comprises a structure model and a tread model. The structure model (Figure 18.21) uses some 1000–5000 lumped-mass nodes, having three translational DOFs each

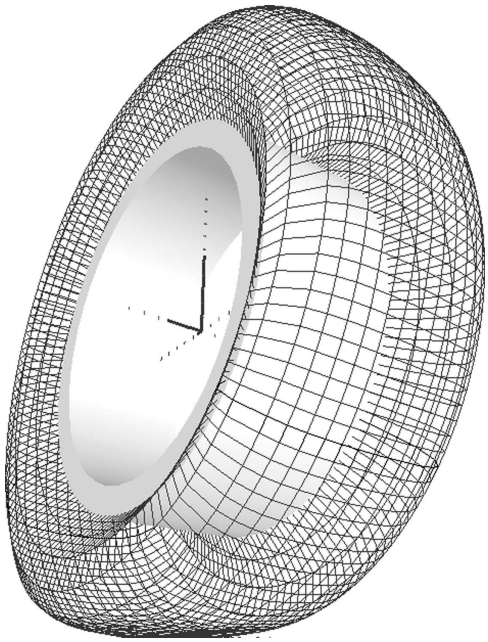


FIGURE 18.21 FETire: mesh of structure model.

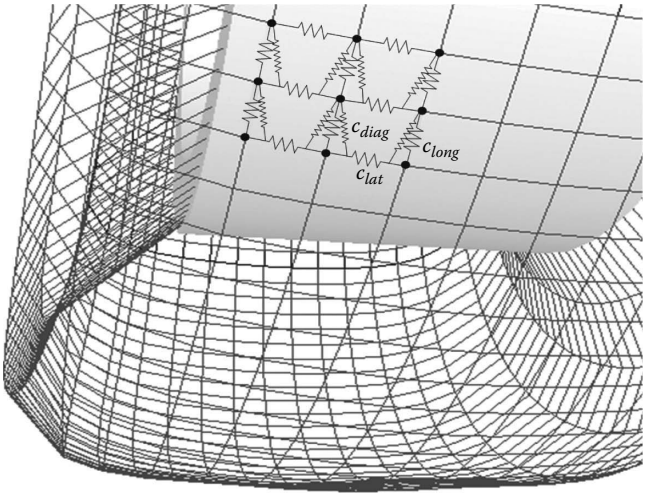


FIGURE 18.22 Springs replacing anisotropic membrane stiffness.

and being connected by a nonlinear network of springs and dampers (Figure 18.22). This coarse FE mesh is automatically generated, using the cross-sectional geometry of the unloaded and uninflated tire only. The distributed membrane stiffness and damping of the layered tire shell structure, consisting of carcass layers, belt layers, bandage layer, bead filler, etc., are computed during preprocessing. It is replaced by the earlier mentioned anisotropic, geometrically fully nonlinear spring and damper network (dampers not displayed in Figure 18.22). Stiffness and orientation of the springs are computed such that the resulting linearized stiffness matrix exactly coincides with the membrane stiffness matrix of the layered structure, separately in each cross-sectional region. By this procedure, for example, very stiff springs are created in the belt region, following the two belt cord direction angles (c_{diag} in Figure 18.22).

This membrane stiffness model is completed by the plate stiffness model, which uses discrete bending stiffness elements. These elements connect any three longitudinally and laterally neighbored nodes in the belt and sidewall region (Figure 18.23).

Finally, radial forces due to inflation pressure contribute to the nodal forces. They act along the shell normal direction, being in turn a nonlinear function of all neighbored node positions. Inflation

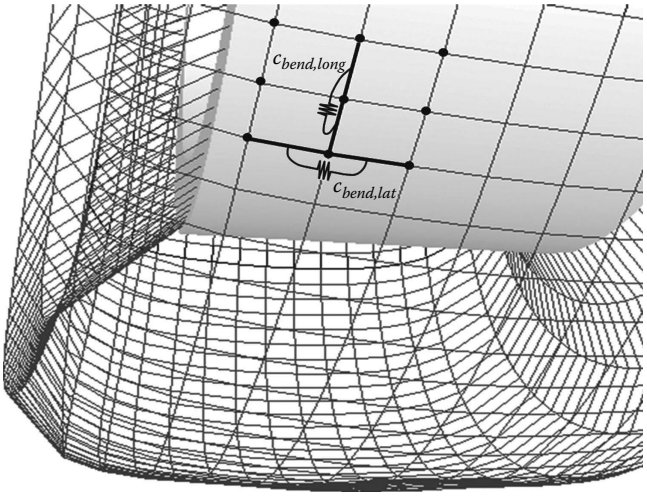


FIGURE 18.23 Longitudinal and lateral bending stiffness in shell structure.

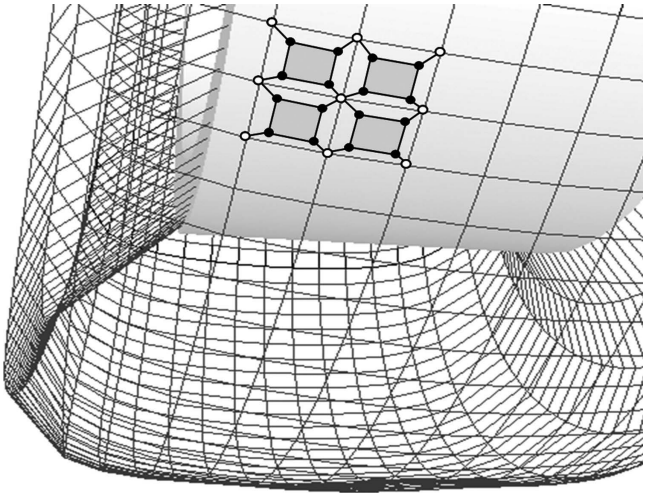


FIGURE 18.24 Tread nodes (●) with associated shell nodes (○).

pressure is computed as thermodynamic function of inner volume and air temperature. Inner volume, in turn, is a simultaneous function of all shell node positions.

Similarly as with FTire, a highly specialized Newmark-type implicit integration is applied to the nodal DOFs. This integrator takes full advantage of the sparsity pattern of the system Jacobian and exploits the fact that the tire structure undergoes largest strains in the vicinity of the contact patch only.

In contrast to FTire, the tread model consists of FEs that are equipped with mass (Figure 18.24). To each quadruple of neighbored shell nodes in the belt region (○), a certain number (4, say) of tread nodes (●) are associated. These tread nodes are coupled by a nonlinear radial and tangential stiffness each to four associated shell nodes. The treatment of road contact and friction is essentially the same as in FTire.

There are two special versions of FETire available. FETire/modal performs linearization and modal analysis of the unloaded tire (Figure 18.25). FETire/static is used for rapid nonlinear static load-case calculation of the loaded but standing tire (Figure 18.26). The latter uses mesh refinement near the contact patch. For rolling tires, this refinement would lead to an unacceptable computational overhead when shell nodes enter or leave the refinement region.

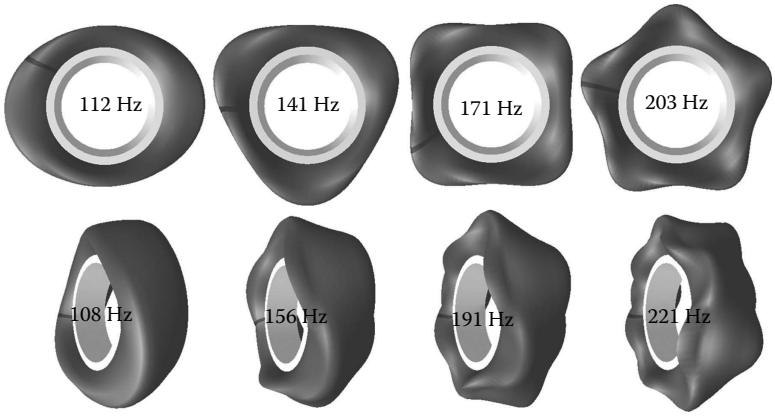


FIGURE 18.25 FETire/modal: first eight in-plane and out-of-plane bending modes of an unloaded passenger car tire. 5130 DOFs, 6 s CPU time for all modes.

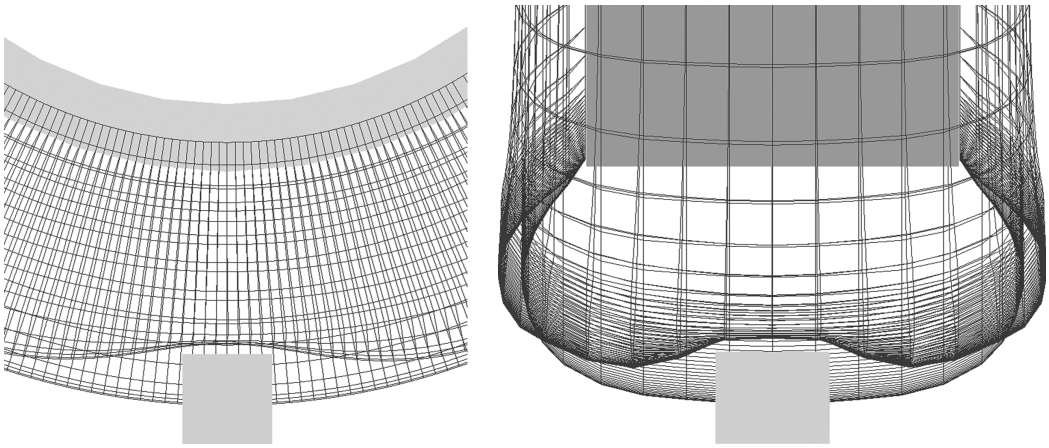


FIGURE 18.26 FETire/static: load-case "transversal cleat" and "longitudinal cleat." Deflection 35 mm, structural distortion compared to that of unloaded tire. 6750 DOFs, 92 s CPU time per load case.

Both special versions are optimized for CPU time and can be used to assist in FTire parameterization (cf. Section 18.7).

To compute stiffness data of the shell structure, FETire uses a parameterized description of the plies and all other important tire compounds. This description is implemented in terms of template files, called tire design data files. Each tire design data file qualitatively describes a certain class of tires, like passenger car tires and motorcycle tires. In such a file, the tire's cross section is separated into several regions, like bead core, apex, sidewall center, shoulder, and belt zenith. Associated to each region, there is a layer list, enumerating each layer together with its thickness, matrix rubber stiffness, cord stiffness (if any), and cord thread angle. Figure 18.27 gives an example of such a layer list.

A layer list does not contain numerical values but rather few variables only. The actual values of these variables, like belt cord density, belt cord direction angle, and layer height, together with a spline-based description of cross-sectional geometry and tread depth, finally are added to the tire design data file to describe a specific tire. To a certain degree, this approach allows a very simple extrapolation, starting from the data of one tire size to estimate the data of another, but similar, one.

Of course, the accuracy of FETire does not reach the one of detailed FE models. FETire is a proven compromise between accuracy and computational efficiency. It is of value as long as detailed FE models cannot yet be used, for many different reasons, in time-domain nonlinear vehicle dynamics simulations.

```
$layer_list_8 ! belt zenith
cr.sect.   cord dens.                              E cord    E matrix
mm^2         thr/mm      height mm    alpha deg    N/mm^2     N/mm^2        type
s_belt = 0.5
0.0          0.0          h_il         0.0          0.0        E_il      ! inner liner
A_carc       d_carc       h_carc       ang_car1     E_carc     E_cp      ! 1. body ply
A_carc       d_carc       h_carc       ang_car2     E_carc     E_cp      ! 2. body ply
A_belt       d_belt       h_belt       ang_belt     E_belt     E_bp      ! 1. belt ply
A_belt       d_belt       h_belt       -ang_belt    E_belt     E_bp      ! 2. belt ply
A_carc       d_band       h_band       0.0          E_carc     E_b       ! 1. bandage
A_carc       d_band       h_band       0.0          E_carc     E_b       ! 2. bandage
0.0          0.0          h_base       0.0          0.0        E_tr      ! tread base
```

FIGURE 18.27 Single layer list of a passenger car tire design data file.

18.7 PARAMETERIZATION TOOLS

Parameterization of any tire model, as mentioned earlier, always is a delicate task. Basically, there are five different procedures conceivable:

1. Direct determination of parameters by using appropriate measurements
2. Usage of a detailed FE model, together with a "condensation process," to directly compute the parameters
3. Usage of measurements of the global tire behavior to estimate the model parameters by means of least squares approximation ("parameter identification")
4. Performance of simulations with a detailed FE model, to generate "virtual measurements," subsequently used for parameter identification as in (3)
5. Usage of a well-validated data file and application of certain "extrapolation formulae" to estimate the tire data of another size, "not too far away" from the validated size

Clearly, due to the different nature of model parameters, none of these five procedures will be used unadulterated. Some data will always be measured directly, for example, simple geometric properties that are only depending on tire size or on tread rubber friction coefficients. Some other data can never be determined directly, for example, a property as artificial as "the stiffness matrix coupling two adjacent belt segments." For the FTire model family, as an example, tools that assist with procedures of types (3), (4), and (5) are available (cf. Figure 18.28).

The first one, called FTire/calc, performs modal analysis using FETire/modal as well as a series of static load-case calculations with FETire/static. These calculations are controlled by a simulation script and run fully automatically. They compute the static tire forces and moments under varying conditions, like

- Vertical deflection on a flat surface
- Vertical deflection on a transversally and a longitudinally oriented cleat
- Vertical deflection with different camber angles on flat surface and on cleats
- Lateral wheel displacement, at a high friction value
- Longitudinal wheel displacement, at a high friction value
- Wheel rotation about vertical axis, at a high friction value

All these computations are performed with several different deflection values and inflation pressures. The resulting static data are completed by the first resonance frequencies and modal damping values, computed with FETire/modal. Here, the same tire design data is used as with FETire/static. Some of the results are used to generate an FTire input file, and some others to validate the respective FTire model. A full cycle of an FTire parameterization with FTire/calc takes less than a minute CPU time.

FTire/fit, the second tool, assists with several procedures of type (3), completed by some handy routines for preparation and preprocessing of experimental data. It provides optimization routines that aim to minimize the least squares distance between measurement data and respective simulations with FTire, for static load cases, for steady-state rolling conditions, as well as for cleat tests. In addition, FTire/fit assists in repeatedly and rapidly performing script-based validation runs and generating detailed result reports.

FTire/fit is not meant to be a "black-box tool," which provides results by few mouse clicks only. Rather, it requires a certain expertise to select and prepare appropriate experimental data and to decide what parameters should be determined by what kind of tests and in what sequence. Moreover, it requires expertise as well to assess whether or not the overall identification result is optimal. As with all applications of computer optimization, there is no guarantee at all that the global optimum really was found. This introduces uncertainty about the quality of the result, which only can

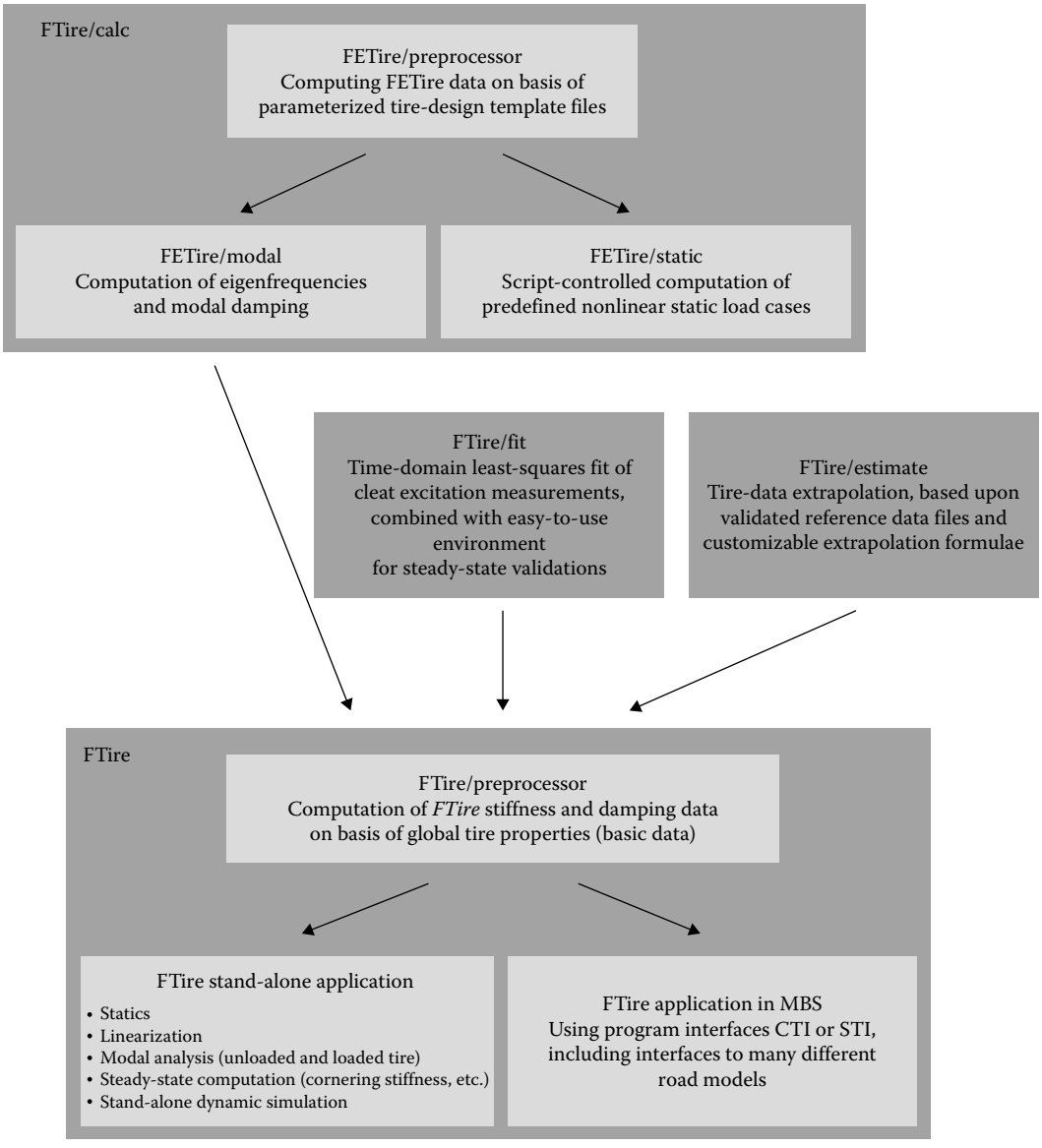


FIGURE 18.28 FTire parameterization tools.

be overcome by having experience with previous similar computations. It seems to be logical to only use a standardized measuring procedure and to use FTire/fit always in the same way in conjunction with this procedure. This is the best way to achieve comparable FTire data for different tires.

Finally, there is a third tool (FTire/estimate, Figure 18.29) available that implements a procedure of type (5). FTire/estimate uses a well-validated FTire data file, representing a whole family of similar tires. Using this tire, it relates its stiffness and modal data to the respective values of other tires in the family, which might differ in section width, aspect ratio, rim size, mass, maximum load, maximum speed, radial stiffness, or nominal inflation pressure.

This relation is described in terms of functional dependencies: The changes in static and modal tire data are functions of the change in one or more influence parameters. These influence parameters are tire section width, tire diameter, sidewall height, tire mass, rated maximum load, rated maximum speed, radial stiffness, and nominal inflation pressure.

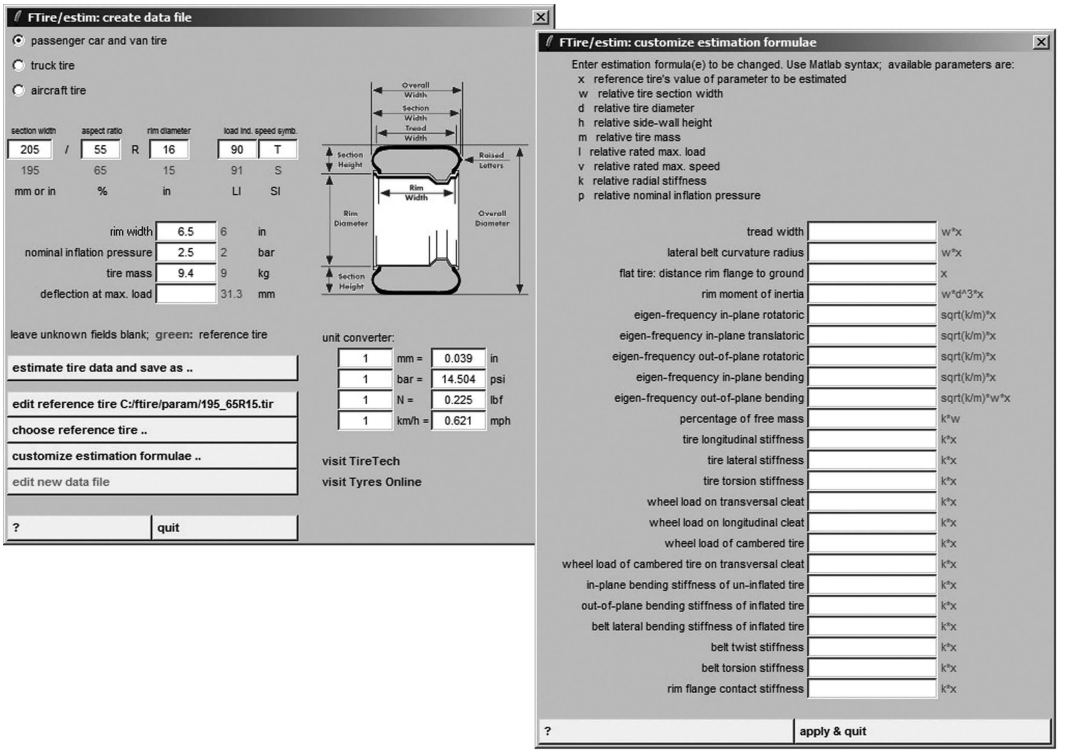


FIGURE 18.29 FTire/estimate user interface (GUI) showing default estimation formulae.

Clearly, these data items are not all independent on each other. The estimation program provides some of them for completeness only. For example, it might be more natural to express the dependency of an eigenvalue on mass than on tire size, even if both will change simultaneously. In any case, it is sufficient to use only the most relevant influence parameters in the formulae.

The influence functions used by default are based upon simple mechanical analogy considerations. However, they can be easily customized. A user might wish to enhance his copy of FTire/estimate, based upon his experience, in the sense of an expert system. Clearly, this experience can be easily completed by studies with detailed FE models or by evaluating respective measurements.

REFERENCES

1. Pacejka H.B., *Tyre and Vehicle Dynamics*, Butterworth and Heinemann, Oxford, U.K., 2012.
2. Kikuchi N., Oden J.T., *Contact Problems in Elasticity*: *A Study of Variational Inequalities and Finite Element Methods*. Vol. 8, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1988.
3. Deneuvy A.C., Modal analysis of a pneumatic tyre and dynamic simulation by a sub-structuring method. *Proceedings of the 7th International Model Analysis Conference*, Las Vegas, NV, 1989.
4. Ishihara K., Development of a three-dimensional membrane element for the finite element analysis of tires. *Tire Science and Technology*, 19(1), 23–36, 1991.
5. Gent A.N., *Engineering with Rubber*. Hanser, München, Germany, 1992.
6. Murakoshi H. et al., An approach to vehicle pull using a tire FE model. *Tire Science and Technology*, 20(4), 212–229, 1992.
7. Le Tallec P., Rahier C., Numerical models of steady rolling for nonlinear viscoelastic structures in finite deformations. *International Journal for Numerical Methods in Engineering*, 37, 1159–1186, 1994.
8. Mark J., Erman B., Eirich F.R. (eds.), *Science and Technology of Rubber*. Academic Press, San Diego, CA, 1994.
9. Abe A., Kamegawa T., Nakajima, Y., Optimum young's modulus distribution in tire design. *Tire Science and Technology*, 24, 204, 1996.

10. Nakajima Y., Kamegawa T., Abe, A., Theory of optimum tire contour and its application. *Tire Science and Technology*, 24, 184, 1996.
11. Meschke G., Payer H.-J., Mang H.A., 3D simulations of automobile tires, material modeling, mesh generation and solution strategies. *Tire Science and Technology*, 25(3), 154–176, 1997.
12. Kennedy R.H., Engelhart M.L., Day, G.L., RPA measurement of hysteresis for CAE rolling resistance prediction. *Proceedings of the International Tire Exhibition and Conference* (*ITEC*), Akron, OH, 1998.
13. Rothert H., Becker A., Dorsch V., Kaliske M., A material model for simulating the hysteretic behavior of filled rubber for rolling tires. *Tire Science and Technology*, 26, 132–148, 1998.
14. Song T-S., Lee J-W., Yu H-J., *Rolling Resistance of Tires—An Analysis of Heat Generation*. SAE International Congress and Exhibition, Detroit, MI, SAE 980255, 1998.
15. Calipel J., *Modélisation numérique du contact*. INRIA Seminar on Applied Nonlinear Problems, Rocquencourt, France, 1999.
16. Nakajima Y., Kadowaki H., Kamegawa T., Ueno K., Application of a neural network for the optimization of tire design. *Tire Science and Technology*, 27, 62, 1999.
17. Dorfi H.R., Wheeler R.L., Griffiths G.H., Tire vibration transmission part I: FEA Eigensolution correlation. *Proceedings of the International Modal Analysis Conference (IMAC)*, San Antonio, TX, 2000.
18. Kennedy R.H., McMinn M.S., Tire temperature prediction during post-cure inflation. *Tire Science and Technology*, 28(4), 248–263, 2000.
19. Okano T., Koishi M., A computational procedure to predict transient hydroplaning of a tire. *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.
20. Seta E., Nakajima Y., Kamegawa T., Ogawa H., Hydroplaning analysis by FEM and FVM: Effect of tire rolling and tire pattern on hydroplaning. *Tire Science and Technology*, 28, 140, 2000.
21. Shiraishi M., Yoshinaga H., Miyori A., Takahashi E., Simulation of dynamically rolling tire. *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.
22. Faßbender F.R., *Simulation der Vertikaldynamik von Fahrzeugen auf Geländeböden mit STINA—Soil Tire Interface to ADAMS, einem Zusatzmodul für das Mehrkörperprogramm ADAMS*. Diss. UniBW Hamburg, Germany, 2001.
23. Yu H.J., Dynamics of tire, wheel & suspension assembly. *Proceedings of the 20th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2001.
24. Kennedy R.H., Prediction of tire shape change during post-cure inflation. *Tire Science and Technology*, 29(4), 198–215, 2001.
25. Koide M., Heguri H., Kamegawa T., Nakajima, Y., Ogawa H., Optimization for motorcycle tire using explicit FEM. *Tire Science and Technology*, 29(4), 2001.
26. Shoop S., Finite element modeling of tire-terrain interaction. PhD thesis, Michigan State University, Ann Arbor, MI, 2001.
27. Tsihlas D., Lacroix T., Clayton B., Comparison of sub-structuring techniques for the dynamic behavior of tires. *Proceedings of the 20th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2001.
28. Darnell I., Mousseau R., Hulbert G., An analytical and experimental investigation into tire force and moment behavior during side slip. *Proceedings of the 21st Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2002.
29. Dorsch V., Becker A., Vossen L., Enhanced rubber friction model for finite element simulations of rolling tyres. *Plastics, Rubber and Composites*, 31(10), 458–464, 2002.
30. Duvernier M., Fraisse P., Bomblain V., Dormegnie E., Tyre modelling for NVH engineering in ADAMS. *Proceedings of the 1st MSC ADAMS European Users' Conference*, London, U.K., 2002.
31. Ohishi K., Suita H., Ishihara K., The finite element approach to predict the plysteer residual cornering force of tires. *Tire Science and Technology*, 30(2), 122–133, 2002.
32. Olatunbosun O.A., Burke A.M., Finite element modeling of rotating tyres in the time domain. *Proceedings of the 21st Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2002.
33. Kennedy R.H., Experiences with cylindrical elements in tire modeling. *Proceedings of the ABAQUS User Conference*, München, Germany, 2003.
34. Nakajima Y., Takahashi F., Abe, A., Surface shape optimization of tire pattern by optimality criteria. *Tire Science and Technology*, 31, 2, 2003.
35. Kennedy R.H., Terziyski J.M., Experiences with embedded elements in tire modeling. *Proceedings of the ABAQUS User Conference*, Boston, MA, 2004.
36. Zhong D., Prediction of tire tread wear with FEM steady state rolling contact simulation. *Tire Science and Technology*, 31(3), 189–202, 2003.

37. Seta E., Kamegawa T., Nakajima Y., Prediction of snow/tire interaction using explicit FEM and FVM. *Tire Science and Technology*, 31(3), 173–188, 2003.
38. Olatunbosun O.A., Bolarinwa O., FE simulation of the effect of tire design parameters on lateral forces and moments. *Tire Science and Technology*, 32(3), 146–163, 2004.
39. Hall W., Mottram J.T., Jones R.P., Tire modeling methodology with the explicit finite element code LS-DYNA. *Tire Science and Technology*, 32(4), 236–261, 2004.
40. Oida S., Seta E., Heguri H., Kato K., Soil/tire interaction analysis using FEM and FVM. *Tire Science and Technology*, 33(1), 38–62, 2005.
41. Bozdog D., Olson W.W., An advanced shell theory based tire model. *Tire Science and Technology*, 33(4), 227–238, 2005.
42. Kabe K., Rachi K., Takahashi N., Kaga Y., Tire design methodology based on safety factor to satisfy tire life (simulation approach to truck and bus tire design). *Tire Science and Technology*, 33(4), 195–209, 2005.
43. Lee C., Rim slip and bead fitment of tires: Analysis and design. *Tire Science and Technology*, 34(1), 38–63, 2006.
44. Shoop S., Kestler K., Haehnel R., Finite element modeling of tires on snow. *Tire Science and Technology*, 34(1), 2–37, 2006.
45. Narasimha Rao K.V., Krishna Kumar R., Bohara P.C., Mukhopadhyay R., A finite element algorithm for the prediction of steady-state temperatures of rolling tires. *Tire Science and Technology*, 34(3), 195–214, 2006.
46. Yin H.S., Hu Y.S., Zhang H., Yang M.M., Wei Y.T., Truck tire thermal-mechanical FEA and DMA with application to endurance evaluation. *Tire Science and Technology*, 34(4), 220–236, 2006.
47. Liu H.H., Load and inflation effects on force and moment of passenger tires using explicit transient dynamics. *Tire Science and Technology*, TSTCA, 35(1), 41–55, 2007.
48. *Tire Science and Technology*. Quarterly published journal, ISSN 0090-8657. The Tire Society, Akron, OH.
49. Gipser M., DNS-Tire, ein dynamisches, räumliches nichtlineares Reifenmodell. In: Reifen, Fahrwerk, Fahrbahn. VDI Berichte 650, pp. 115–135. VDI Verlag, Düsseldorf, North Rhine-Westphalia, Germany, 1987.
50. Gipser M., DNS-tire, a dynamic nonlinear spatial tire model in vehicle dynamics. *Proceedings of the 2nd Workshop on Road-Vehicle-Systems and Related Mathematics*, ISI, Torino, Italy, 1987, Neunzert H.T. (ed.), Stuttgart, Baden-Württemberg, Germany, 1989.
51. Böhm F., *Reifenmodell für hochfrequente Rollvorgänge auf kurzwelligen Fahrbahnen*. VDI Berichte 1088, pp. 65–81. VDI Verlag, Düsseldorf, North Rhine-Westphalia, Germany, 1993.
52. Negrut D., Freeman J.S., Dynamic tire modelling for application with vehicle simulations incorporating terrain. SAE Technical Papers 940223, 1994.
53. Ammon D., Gipser M., Rauh J., Wimmer J., Effiziente Simulation der Gesamtsystemdynamik Reifen-Achse-Fahrwerk. In: Reifen, Fahrwerk, Fahrbahn. VDI-Berichte 1224, VDI Verlag, Düsseldorf, North Rhine-Westphalia, Germany, 1995.
54. Fujioka T., Goda K., Discrete brush tire model for calculating tire forces. Supplement to *Vehicle System Dynamics*, 25, 200–216. Swets & Zeitlinger, Lisse, the Netherlands, 1996.
55. Gipser M., DNS-Tire 3.0—Die Weiterentwicklung eines bewährten strukturmechanischen Reifenmodells. In: *Darmstädter Reifenkolloquium*, VDI Berichte 512, pp. 52–62. VDI Verlag, Düsseldorf, North Rhine-Westphalia, Germany, 1996.
56. Mousseau C.W., Hulbert G.M., Clark S.K., On the modeling of tires for the prediction of automotive durability loads. Supplement to *Vehicle System Dynamics*, 25, 466–488. Swets & Zeitlinger, Lisse, the Netherlands, 1996.
57. Vinesse E.P., Sub-structuring method and eigen function representation for a rolling tyre coupled to a secondary mechanical system. *Proceedings of the International Mechanical Engineering Congress and Exhibition*, Atlanta 1996.
58. Darnell I., Hulbert G.M., Mousseau C.W., An efficient three-dimensional tire model for vehicle dynamics simulation. *Mechanics of Structures and Machines*, 25(1), 1–19, 1997.
59. Gipser M., Hofer R., Lugner P., Dynamical tire forces response to road unevenness. *Proceedings of the 2nd International Colloquium on Tyre Models for Vehicle Dynamic Analysis*. Supplement to *Vehicle System Dynamics*, Vol. 27. Swets & Zeitlinger, Lisse, the Netherlands, 1997.
60. Mastinu G., Gaiazzi S., Montanaro F., Pirola D., A semi-analytical tyre model for steady—and transient-state simulations. *Proceedings of the 2nd International Colloquium on Tyre Models for Vehicle Dynamic Analysis*. Supplement to *Vehicle System Dynamics*, Vol. 27. Swets & Zeitlinger, Lisse, the Netherlands, 1997.

61. Oertel C., On modelling contact and friction: Calculation of tyre response on uneven roads. Supplement to *Vehicle System Dynamics*, 27, 289–302. Swets & Zeitlinger, Lisse, the Netherlands, 1997.
62. Mancosu F., Da Re D., Minen D., Non-linear modal rolling tyre model for dynamic simulation with ADAMS. *Proceedings of the 13th European ADAMS Users' Conference*, Paris, France, 1998.
63. Gipser M., Reifenmodelle für Komfort—Und Schlechtwegsimulationen. *Proceedings of the 7th Aachener Kolloquium Fahrzeug—Und Motorentechnik*, IKA, RWTH, Aachen, North Rhine-Westphalia, Germany, 1998.
64. Fandre A., Oertel C., Ride comfort simulations and steps towards life time calculations: RMOD-K tyre model and ADAMS. *Proceedings of the 14th European ADAMS Users' Conference*, Berlin, Germany, 1999.
65. Ferrarese A., Padovese L.R., Costa A.L.A., Tire dynamical model. *Computational Methods in Engineering*, Vol. 195, 1–12, 1999.
66. Zhou J., Wong J.Y., Sharp R.S., A multi-spoke, three plane tyre model for simulation of transient behaviour. Supplement to *Vehicle System Dynamics*, 31. Swets & Zeitlinger, Lisse, the Netherlands, 1999.
67. Kao B., A three-dimensional tire model for vehicle dynamic simulations. *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.
68. Mancosu F., Sangalli R., Chel F., Ciarlariello G., Braghin F., Comparison of a new mathematical-physical 2D tyre model for handling optimization on a vehicle with experimental results. *Proceedings of the 19th Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2000.
69. Deur J., Asgari J., Hrovat D., A dynamic tire friction model for combined longitudinal and lateral motion. *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE)*, Vol. 2, New York, 2001.
70. Mancosu F., Savi C., Brivio P., Travaglio G., Ramirez I., New dynamic tyre model in multi-body environment. SAE 2001 World Congress, Detroit, MI, SAE Technical Paper Series 2001–01–0747, 2001.
71. Oertel C., Eigensystem of tangential contact in tyre models. *Vehicle System Dynamics*, 31, 232–241, 2001.
72. Bäcker M.H., Köttgen V.B., Dressler K., LMS CDTire model: Road load simulation on a digital test track. *LMS Conference US*, Troy, OH, 2002.
73. Berzeri M., Dhir A., Ranganathan R., Balendran B., Jayakumar P., O'Heron P., A new tire model for road loads simulation: Part 1: Theory and validation. *Proceedings of the 22nd Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2003.
74. Wei Y.T., Nasdala L., Rothert H., Analysis of tire rolling contact response by REF model. *Proceedings of the 22nd Annual Meeting and Conference on Tire Science and Technology*, Akron, OH, 2003.
75. Gipser M., FTire, a new fast tire model for ride comfort simulations. *International ADAMS User's Conference*, Berlin, Germany, 1999.
76. Van Oosten J.J.M., Unrau H.J., Riedel G., Bakker E., *TYDEX Workshop: Standardisation of Data Exchange in Tyre Testing and Tyre Modelling*. In: *Tyre Models for Vehicle Dynamic Analysis*. Supplement to *Vehicle System Dynamics*, 27. Swets & Zeitlinger, Lisse, the Netherlands, 1997.
77. Gipser M., *FTire Documentation* (http://www.cosin.eu/res/ftire_model.pdf). Accessed May 13, 2013.
78. Gipser M., ADAMS/FTire—A tire model for ride & durability simulations. *ADAMS User's Conference*, Tokyo, Japan, 2000.
79. Riepl A., Reinalter W., Fruhmann G., Rough road simulation with tire model RMOD-K and FTire. *Proceedings of the 18th IAVSD Symposium*, Kanagawa Institute of Technology, Atsugi-shi, Japan, 2003.
80. Dorfi H.R., A study of the in-plane force transmission of tires. *Tire Science and Technology*, 32(4), 188–213, 2004.
81. Dorfi H.R., Tire non-uniformities and steering wheel vibrations. *Tire Science and Technology*, 33(2), 64–102, 2005.
82. Gipser M., FTire software: Advances in modelization and data supply. *Proceedings of the Tire Society Meeting*, Akron, OH, 2006.
83. Gipser M., Reifensimulation mit FTire: Stand und Ausblick. 15. *Aachener Fahrzeug - und Motorenkolloquium*, Aachen, Germany, 2006.
84. Gipser M., *FTire: 10 Years of Development and Application. Vehicle System Dynamics*, 45, Supplement, Taylor & Francis Group, Oxford, 2008.

19 Pneumatic Tire
Construction and Testing

Maurizio Boiocchi and Giuseppe Matrascia

CONTENTS

19.1 Introduction 658
19.2 Materials 658
19.2.1 Rubber 659
19.2.1.1 Natural Rubber 659
19.2.1.2 Synthetic Rubber 659
19.2.2 Fillers 660
19.2.2.1 Black Fillers 660
19.2.2.2 Nonblack Fillers 660
19.2.3 Curing System 660
19.2.4 Protective System 661
19.2.5 Processing Agents 661
19.2.6 Resins 662
19.2.7 Steel and Textile 662
19.2.7.1 Steel 662
19.2.7.2 Textile 663
19.3 Manufacturing Process 663
19.3.1 Mixing 663
19.3.2 Calendering/Extrusion 664
19.3.2.1 Calendering 664
19.3.2.2 Extrusion 664
19.3.3 Building 668
19.3.3.1 First Stage 668
19.3.3.2 Second Stage (Conformation) 668
19.3.4 Curing 669
19.3.4.1 Introduction to Curing 669
19.3.4.2 Curing Cycle 670
19.3.4.3 Curing Mold 670
19.3.4.4 Lubricant for Curing Cycle 670
19.3.4.5 Spews 672
19.4 Tire Structure 672
19.4.1 Inner Liner 673
19.4.2 Plies 673
19.4.3 Bead and Filler 674
19.4.4 Chafer and Rim Cushion 674
19.4.5 Sidewall 675
19.4.6 Belts 675
19.4.6.1 Nylon 0° 675

19.4.7 Belt Wedge: Edge Cover 676
19.4.8 Tread 676
19.5 Tire Standards and Tests 677
19.5.1 UN/ECE Regulations 677
19.5.1.1 Markings 678
19.5.1.2 Tire Dimensions 678
19.5.1.3 Load and Speed Performance Test 682
19.5.1.4 Approval of the Tire 685
19.5.2 DOT Tests FMVSS139 685
19.5.2.1 Tire Dimensions 685
19.5.2.2 Bead Unseating Test 685
19.5.2.3 Tire Strength 685
19.5.2.4 Tire Endurance 687
19.5.2.5 Low-Pressure Test 688
19.5.2.6 High-Speed Performance 688
19.5.3 Internal Experimental Tests 688
19.5.3.1 Type A Tests 688
19.5.3.2 Type B Tests 688
19.5.4 Other Standards or Regulations 694
19.6 Conclusions 694
References 694

19.1 INTRODUCTION

The pneumatic tire ("aerial wheel") was conceived by R.W. Thompson (1845) and then reinvented by J. Dunlop (1888) as a composite material structure. More than 200 different raw materials are used to build a modern pneumatic tire. The pneumatic tire (briefly called tire in the sequel) is truly one of the very first successful applications in the field of composite materials [1].

The mechanics of pneumatic tire presented in Chapters 17 and 18 can be understood better if proper attention is paid to the chemical and physical properties of the materials that the tire is made from. For such a better understanding, also the construction process, the details on tire inner structure, and the safety assessment tests are to be taken into account.

The specific literature on tire materials, structure, and manufacturing is relatively limited, the up-to-date information can be found on the Internet; this is the reason why many references to web pages have been introduced at the end of the chapter [2].

In the chapter, the radial pneumatic tire will be considered [3]. The characteristics of materials, manufacture, and testing of other tire types (bias, diagonal) are not much different from the ones of the radial tire.

In the first part of the chapter, the materials used to build a pneumatic tire are presented. In the second part of the chapter, the manufacturing process is introduced. In the third part of the chapter, the tire's inner structure is analyzed. In the fourth and last part of the chapter, the tests that are performed to deliver a tire are listed and commented.

19.2 MATERIALS

The pneumatic tire is made from many components, both reinforcement materials and compounds (Figure 19.16). Compounds are specific blends of materials developed to achieve the specific performance required. The main families of materials in the pneumatic tire are

- Rubbers
- Fillers (carbon black, silica)
- Curing system (sulfur, accelerators)

TABLE 19.1
Typical Material Composition of a Pneumatic Tire (Reinforcement Excluded)

Component	%	Type
Rubber	50.0	Rubber
Carbon black	17.5	Filler
Silica	17.5	Filler
Oil	10.0	Processing aid
Resins	2.0	Enhancers of different properties
Antioxidant	1.0	Protective system
Wax	1.0	Protective system
Zinc oxide	2.0	Activator of the curing system
Stearic acid	0.8	Activator of the curing system
Accelerator	0.5	Curing system
Sulfur	0.8	Curing system

- Protective system (antioxidant, antiozone, wax)
- Processing aids
- Steel and textile cords

A typical material composition of a tire could be the one reported in Table 19.1 (the reinforcement cords are excluded). Obviously, before and during the tire manufacture [4], special control tests are performed in order to monitor all of the parameters that characterize each material of the compound and the compound itself.

19.2.1 Rubber

There are many kinds of rubber either natural or synthetic involved in the manufacture of the tire [3,5].

19.2.1.1 Natural Rubber

The natural rubber (NR) is obtained from the sap-like extract (latex) of *Hevea brasiliensis* (Pará), a tall tree that is grown in zones near the earth equator (latitude ±20°) where the rainfall rate is about 1500–2000 mm/year. Such zones are Amazon rainforest, Central America, Central Africa near the Atlantic Ocean, India, and Southeast Asia. Natural rubber is mostly composed by *1,4-cis-polyisoprene* (C_5H_8 is the monomer). Natural rubber was scientifically discovered by C.M. de La Condamine in 1745, although C. Columbus noted rubber balls used by native Americans (rubber was used since 1600 BC by Mesoamerican people). Natural rubber resists more to stress and has less hysteresis than synthetic rubbers. Several types of natural rubber are identified and used [1].

19.2.1.2 Synthetic Rubber

Since about 1945 synthetic rubber was employed for tire manufacturing. *International Institute of Synthetic Rubber Producers (IISRP)* provides the classification of synthetic rubbers [6].

The more used synthetic rubber is *styrene-butadiene* (SBR), which has a good hysteresis that implies a good grip. *Polybutadiene* (BR) has a good fatigue and wear resistance. *Butyl* rubber, and *halobutyl* in particular being impermeable to gases, is used in the inner liner of the tire to prevent air leakage. Obviously, the rubbers have to be mixed with the other components in order to obtain a balance of desired material properties.

Mechanical properties of rubbers are greatly influenced by the *glass transition temperature* T_g, which identifies primarily the viscous behavior of the material. Information on T_g can be found in [7].

19.2.2 Fillers

Fillers are very important parts of the tire rubber compounds, because they offer abrasion resistance and reinforcement. Fillers are solid granular materials.

19.2.2.1 Black Fillers

Carbon black (obtained from the combustion of natural gas or hydrocarbons) is the more used filler. Carbon black was added to rubber since about 1915. Carbon black is made of particles of fine size. Its use is devoted to improve fundamental properties such as mechanical resistance, elastic modulus, tear strength, abrasion resistance, UV rays resistance, and hardness.

The technical specifications of carbon black [8–11] include structure, particle size, particle porosity, and overall physical–chemical nature of particle surface. These factors strongly influence the degree of reinforcement and also cure rate [3] (for cure rate and curing, see Section 19.2.3). There are several grades of black fillers ([1] ASTM 1765).

According to the production cycle, the carbon black can have different pH: acid black fillers tend to retard the curing process. On the contrary, alkaline black fillers (made by furnace processing) produce a rate-enhancing effect on curing. If black filler acidity is wrongly tuned, it can give rise to scorching (i.e., vulcanization starts too early).

The particle size of the filler has a primary role. The smaller the particle size the higher the reinforcement, but the poorer the process efficiency because of the longer time needed for dispersion and the greater the heat produced during mixing. Such an occurrence can be explained thinking of black filler that is dispersed into the voids among rubber chains. The finer the black filler, the better the positioning into the voids (and more difficult the dispersion). The finer the dispersion, the higher the modulus of elasticity of the rubber.

19.2.2.2 Nonblack Fillers

The most powerful nonblack filler involved in tire manufacture is silica, because of its small particle size and complex aggregate structure. Silica precipitates impart the highest degree of reinforcement to rubber among all the nonblack particulate fillers. Silica is employed for manufacturing the tire tread, due to wet grip improvement and low rolling resistance. Additionally for off-road tire, it has proved to guarantee the reduction of chucking (a special type of damage of the tread in which small pieces of compound are removed) and tear resistance.

The silica (obtained from special sand) is more expensive than carbon black but produces less rolling resistance and higher grip at low temperature [3,5].

The reason why silica was never used as the primary filler was a lack of compatibility between silica and the other polymers used in the tread (actually special bonding agents—silane—are to be used); in addition silica does not disperse very well during the mixing process. The silane coupled with silica is also useful in improving the abrasion resistance of silicas up to levels of carbon blacks and promotes their use in the treads.

Silica can also be used in blend with carbon black.

19.2.3 Curing System

One of the most essential components of any tire compound is sulfur, which allows the rubber to be vulcanized. Vulcanization was (accidentally) discovered by C. Goodyear in 1839. During vulcanization the sulfur creates cross-links between the carbon backbones of polymer chains when the compound is heated. Cross-linking increases the strength of the rubber by preventing polymer chains from slipping one another. In order to speed up the vulcanization process, ingredients called accelerators are used together with sulfur; the most common accelerators are sulfenamides. Zinc oxide and stearic acid are also used as activators of the curing system: they react with the accelerator so that a "sulfurating complex" is formed. The "sulfurating complex" creates the cross-link network [1].

Figure 19.1 shows a schematic representation of polymer chains before and after vulcanization.

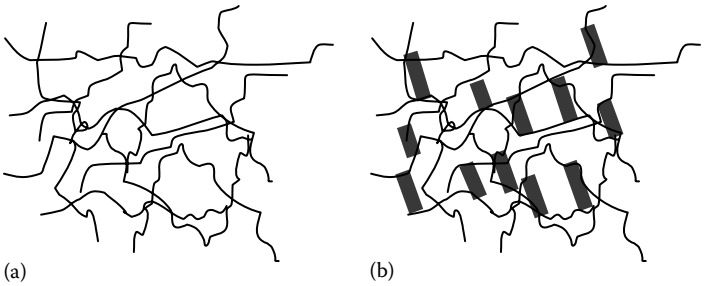


FIGURE 19.1 Schematic representation of polymer chains before (a) and after (b) vulcanization. During vulcanization, sulfur creates cross-links between the carbon backbones.

Sulfur tends to form chains composed by its own atoms (as carbon and silicon do). The cure sites on the rubber molecule can be defined as the sites that are attractive to sulfur atoms. At each cure site on the rubber molecule, a sulfur atom can attach itself and from there a sulfur chain can grow until it reaches a cure site on another rubber molecule. The sulfur bridges can be from 1 up to 10 atoms long, differently from typical polymer in which the carbon backbone is many thousands of atomic unit in length.

19.2.4 Protective System

There are several indices that describe the service life of the rubber products: one of these is the aging index [12–14]. Practically, the aged rubber is more brittle than the nonaged one.

The kind of elastomer is the main factor for aging property. The more the backbone of the elastomer is saturated, the better the aging properties can be expected. Actually, the environment brings the tire in contact with air, and the ozone concentration in the atmosphere rapidly reacts with the carbon elastomer mainly in the backbone unsaturated zones. In general, unsaturated polymers are susceptible to thermal or ozone oxidation. Thus antioxidants and antiozonant are needed to extend the life of the product. Heat and relatively high temperature enhance oxidation. A rise in working temperature of 10° doubles the rate of the unwanted process. Flex cracking is a common mode of fatigue failure in the rubber compound. Flex cracking involves not only mechanical fatigue but also oxidation, which is accelerated by heat generated during the flexure cycle. The higher the rate of flexure, the greater is the heat generation and the worse is the fatigue life of the rubber compound. Conversely, oxidation reaction is accelerated by stress, stretching, and bending of the elastomer. There is a critical elongation below which cracks do not form. Such a critical elongation is 5%–10% of maximum material elongation.

The chemical reaction of rubber with ozone is very fast. Due to this short time, the ozone cannot diffuse inside the tire and remains on the surface. The antiozonant protectors initially sacrifice themselves reacting with ozone, but this reaction produces a protective surface layer that avoids the ozone to continue deteriorating the tire surface. Paraffinic waxes are also widely used as protective agents: they bloom on the surface of the compound to form an inert film.

UV rays deteriorate rubber. In the compound, some anti-UV ingredients are commonly added; they can be simply carbon black or UV stabilizers. The UV light reacts with the surface creating small cracks that form a typical pattern called *crazing*. The ions of iron and copper (coming from steel belt and bead wire) and the ions of manganese and cobalt (in the compound) even if present in small amounts give a great contribution to rubber oxidation. They affect the breakdown of peroxides, so they produce a further acceleration in the oxygen attack. The most sensitive rubber to this phenomena is the vulcanized natural rubber.

19.2.5 Processing Agents

Natural rubber has to be subjected to mastication, and the efficiency of such a power-consuming operation can be increased by adding *peptizers*. Upon a certain temperature, peptizers fasten the

rate of breakdown of rubber chains (common peptizers are thiophenols and diphenyl-disulfides). Peptizers are used to make the rubber soft and tacky, to facilitate mixing operation, especially if a big amount of carbon black is added.

Many rubber types are involved in the compound, and they need to have the same mechanical behavior during the mixing process. To fulfill such a condition, different rubbers have to adapt themselves to the processing cycles. To facilitate such an adaptation, specific *plasticizers* are added. A great influence of such ingredients can be seen on the diene-hydrocarbon rubbers (SBR, IR, BR); for such materials hydrocarbon oils are preferably used. These oils (generally known as softeners) decrease the polymer viscosity and reduce hardness and low-temperature brittle point of the cured product. Three main types of softeners exist: aliphatic, aromatic, and naphthenic. The aromatic oils have been banned and replaced by grades with lower content of PAH (polycyclic aromatic hydrocarbons); the naphthenic grades are also widely used.

Natural rubbers exhibit the phenomena of tack (i.e., when two clean surfaces of masticated rubber are brought in contact, the two surfaces strongly adhere one to each other). Synthetic rubbers in general show low tackiness, so that it is necessary to add *tackifiers* (hydrocarbon resins are widely used). A proper tack is essential during the tire manufacturing.

19.2.6 Resins

Different families of resins are used in compounds for tires. Resins used in the tread compound are mainly grip enhancers: they are essential for increasing hysteresis while keeping the elastic modulus unchanged. Phenolic resins are often used in compounds for bead filler, carcass, and belts.

19.2.7 Steel and Textile

19.2.7.1 Steel

Steel cords are the fundamental structural components of tire. Typical steel involved is the ASTM 1070 steel tire wire; a usual value for tensile strength is 2750 MPa. There are two kinds of steels, one for the belts and one for the bead; their chemical compositions are reported in Table 19.2.

The two steels in Table 19.2 are similar, and little difference on the carbon content can be found.

The steel wire is produced by wire drawing, a cold deformation process. The steel wire passes through a series of dies. The traction is produced by *capstans*. Starting from a rod of 5.5 mm diameter,

TABLE 19.2
Average Chemical Composition of Steel for Belts and Beads

	Steel Belts	Bead Wire
Carbon	0.67–0.73%	0.60% min.
Manganese	0.40–0.70%	0.40–0.70%
Silicon	0.15–0.03%	0.15–0.30%
Phosphorus	0.03% max.	0.04% max.
Sulfur	0.03% max.	0.04% max.
Copper	Trace	Trace
Chromium	Trace	Trace
Nickel	Trace	Trace
Coating	66% Copper	98% Brass
	34% Zinc	2% Tin

wires are obtained whose diameter is a few hundredths of a millimeter. As the cord is 1.2 mm of diameter, it is brass plated and then further reduced in diameter.

Two kinds of wire drawing processes exist, dry or in bath. They differ either for the kind of lubricant or the way the capstan acts. A maximum of wire deformation is accepted for each die; the wire size has to be reduced by means of a series of passages in order to avoid nano-defects inside the metal material lattice. Also microscopic and macroscopic defects like the Chevron creek or broken point (which require welding) are not accepted by tire manufactures. The more the steps of wire deformation, the more the ultimate tensile stress and the less the elongation at rupture.

The wires have to be submitted to brass coating to let rubber adhere firmly to steel. Actually rubber does not bond spontaneously to steel. Brass (copper more exactly) and sulfide create protuberances on the steel wire that help rubber adhere to the wire itself.

Adhesion can sometimes be degraded if copper sulfide is formed on the rubber joining the wire. This implies a weak brittle bound with the underlying brass-plated steel cord and is prejudicial to good adhesion. To avoid such problem, the proper concentration of ingredients in the compound must be used; concentration of copper in brass, as well as the thickness of brass coating, is also crucial to get a good adhesion. Wires are used to produce cords. For instance, a $3 + 6 \times 0.22$ cord is composed by three filaments warped with additional six wires with a diameter of 0.22 mm [1].

19.2.7.2 Textile

Polyester (PET) and rayon are used primarily for carcass plies (see Figure 19.16). Nylon-6, nylon-66, and aramid (aromatic polyamide) are used mainly for 0° ply (see Figure 19.16).

PET is used for passenger, light truck, or farm tires. Nylon-66 is preferred to nylon-6 due to higher melting temperature. Nylon has a good fatigue resistance, tenacity, and durability. Unfortunately, it has low dimensional stability.

Aramid is used as a metal substitute in tires that have to be as light as possible (for racing applications, aramid can be used for belts; see Figure 19.16).

19.3 MANUFACTURING PROCESS

The manufacturing process of a tire is based on

- Mixing
- Calendering/extrusion
- Building
- Curing

19.3.1 Mixing

Mixing the rubber compound is the first phase of the tire manufacture. The Banbury machine is used for mixing. The mixer consists of two counterrotating spiral-shaped blades sheathed in segments of cylindrical shape (Figure 19.2). The movement of the blades leaves a space for the compound between them. The blades are usually made to let a fluid flow inside them; the fluid can be both cold and hot, depending on the material involved. The body and structure of this machine are strong, and they let be suitable heavy-duty mixing. The structure is made of a feed hooper, with a large cross-sectional area for rapid feeding. Rotors are usually made from alloy steel, for mechanical property and capability of productivity. Through the door for discharge control in temperature can be done. Finally, there are a series of seals to avoid dust and there is a cooling system. In Figure 19.3, it is possible to see the sheet of compound leaving the mixer. In Figure 19.4, the parts of interest of the mixer are explained, and three different types of mixer are shown.

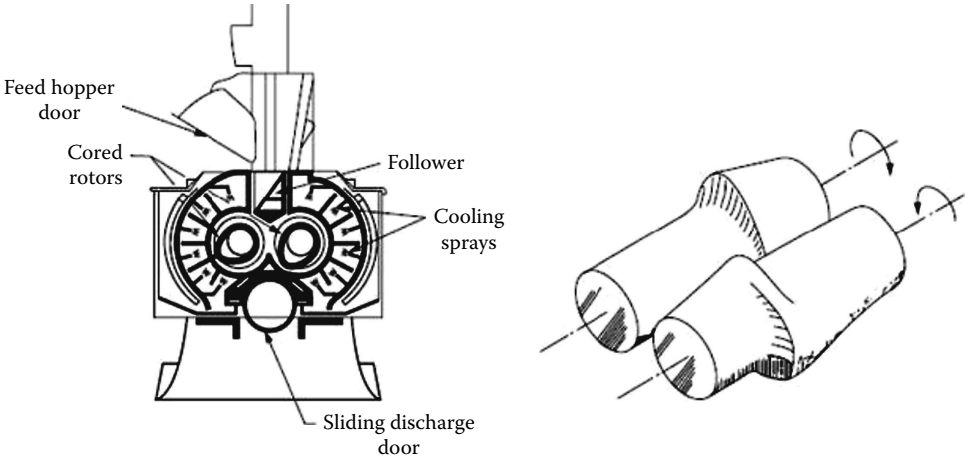


FIGURE 19.2 Banbury chamber and rotors.



FIGURE 19.3 Compound from Banbury.

19.3.2 Calendering/Extrusion

19.3.2.1 Calendering

One fundamental process involved in the tire manufacture is calendering. Calendering is a continuous process and it is used to produce compound sheets of several thickness. Calendering is also used to rubberize cords for plies and belts: both metallic and textile cords are covered with appropriate rubber compounds to get fabrics.

The feeding compound is usually heated, to achieve a viscosity suitable for the process. Usually when the sheet leaves the rolls, it has to be cooled and then rolled on a drum.

The fabric is created by inserting metallic/textile cords between two sheets of compound passing through two or more rolls (see Figure 19.5).

Many different arrangements for calender lines exist, having different numbers and positions of the rolling mills.

Figure 19.6 shows a typical scheme of a calender for metallic fabrics.

19.3.2.2 Extrusion

An energy-demanding operation in the tire manufacturing is extrusion. The extrusion process is designed to continuously convert a feeded material into a desired shape, by means of a die.

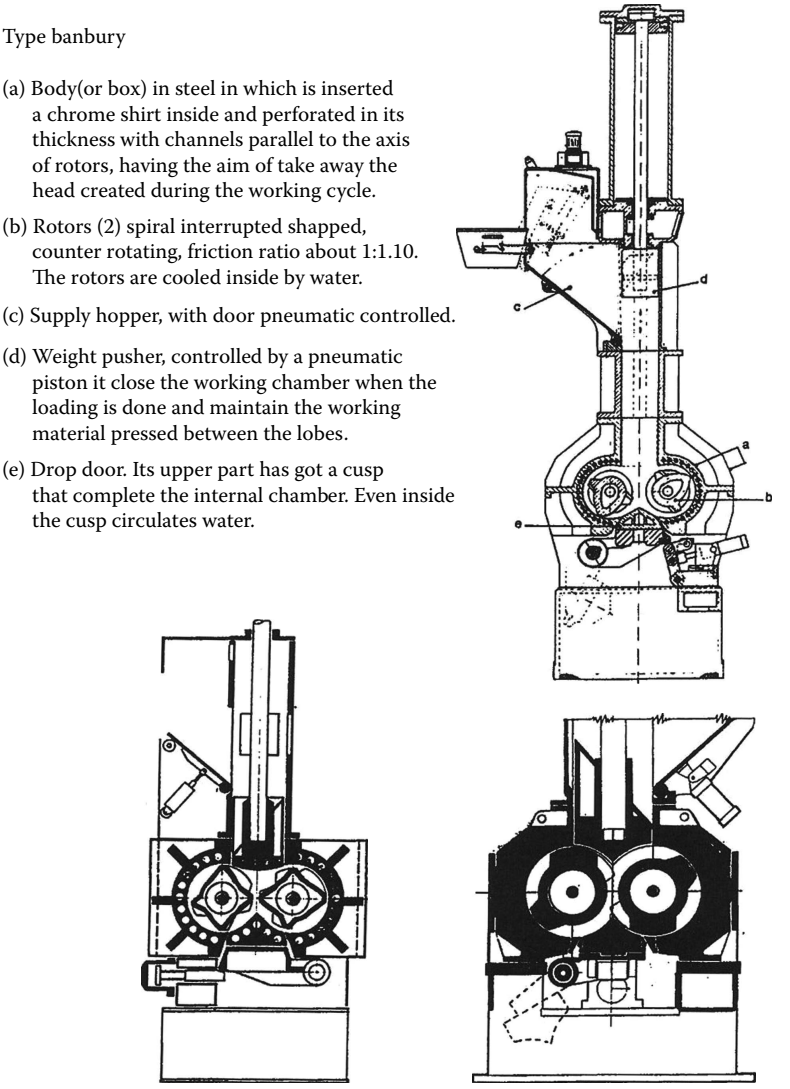


FIGURE 19.4 Banbury, Werner, and Intermix machines.

The material is carried out by the action of a screw, heated till it is squeezed through a die (Figure 19.7). Such operation is mainly used on tread and sidewall–rim cushion compounds.

The environment can contain even the external heat generators. The pressure on the material grows up with the decrease of the distance from the feeder up to the die and reaches the maximum just before it enters into the die. The pressure on the compound forces it through an adapter and then into the die. The die is a shaped open one, to give the extruded compound the desired shape (Figure 19.8).

The control of the process is very important, and the exit temperature of the extruded compound must be in a specific range in order to avoid early vulcanization.

If the temperature have to be reduced, it is possible to act by means of a cooler. The most important component of any extruder is the screw (Figure 19.9). It is proper of one material and cannot

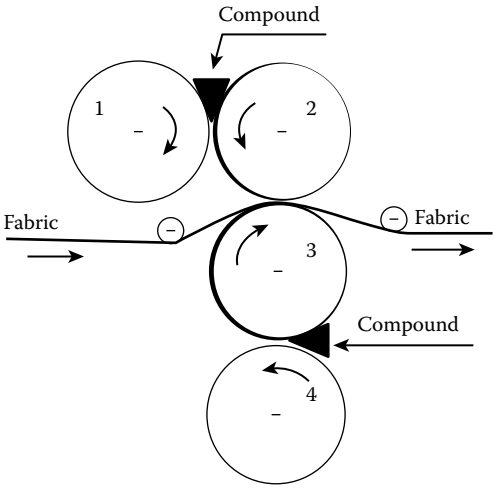


FIGURE 19.5 Zoom on calendering process.

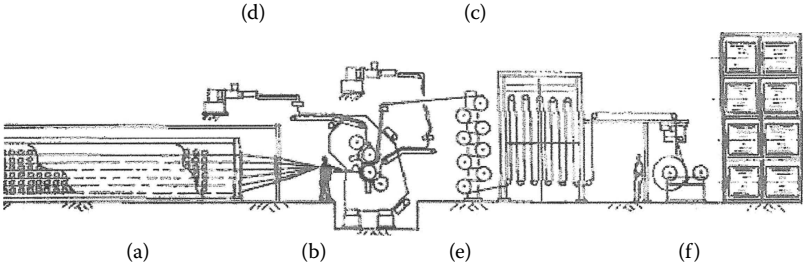


FIGURE 19.6 Calendering machine. (a) Creel: planarity and coupling rubber/steel. (b) Density rollers: uniformity of density, penetration compound, and thickness symmetry. (c) Calandra: thickness (transversal/symmetric/total), weight, and penetration compound. (d) Compound processing: adhesiveness, thickness, and penetration compound. (e) Cooling and tension: planarity flatness and adhesiveness. (f) Gathering and storing of fabric drum: squaring, adhesiveness, and weldability.

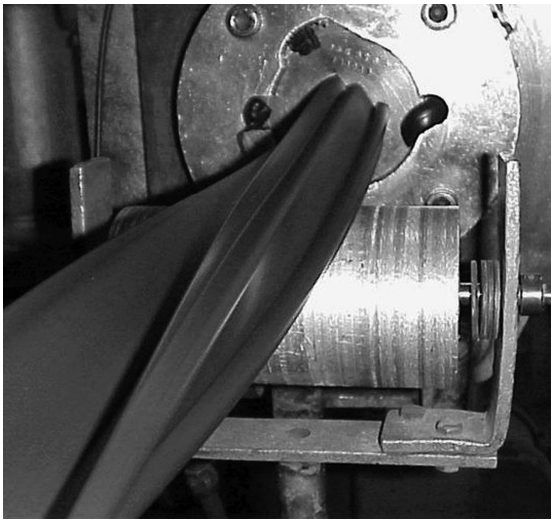


FIGURE 19.7 Compound from extrusion.

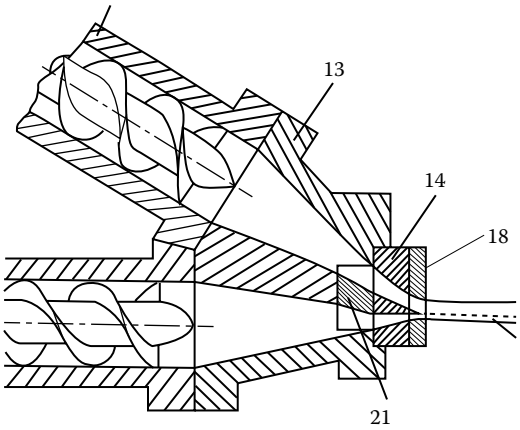


FIGURE 19.8 Double compound extrusion die.

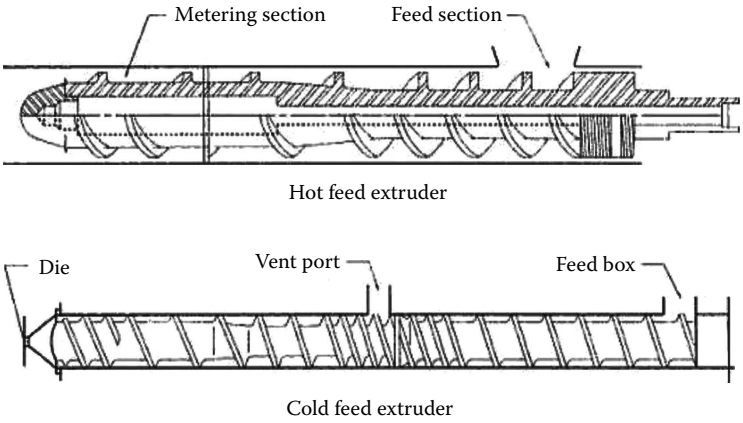


FIGURE 19.9 Cold feed and hot feed extruders.

be used for others. The screws are often fabricated with a 4140 alloy steel in cylindrical shape with a helical channel cut into it. The helical ridge formed by this machining is called the flight, the distance between the flights is called lead, and it can vary along the screw axis. The helix angle is called pitch, and this is properly chosen in order to optimize the feeding characteristics; an angle of 17.6° can be considered as an average, and it usually varies between 12° and 20.1°. The screw outside diameter is generally a few hundredth of millimeter less than the internal diameter of the vessel. This little clearance avoids high steady increase of the compound inside the barrel wall and brings maximization in the heat transfer. The screw flights are usually submitted to heat treatment, in order to increase the hardness, growing up the wearing resistance.

The feeding system can be "hot" or "cold" (Figure 19.10). In the hot feed system, strips of green compound are fed from warm-up mills into an extruder feed box and then into the vessel. This system normally has got short barrel and screws with short compression and metering zones, because the compound is already hot and has a low viscosity as it enters to feed the hoopers; barrel length to diameter ratio from 4:1 to 6:1 and 8:1 is typical. Cold feed systems feed the rubber compound directly into the extruder system, and they have much larger length to diameter ratio because of the requirements of reducing green compound plasticity.

The compound from the extruder machine can be made by more die head to form a sheet with more than one compound.

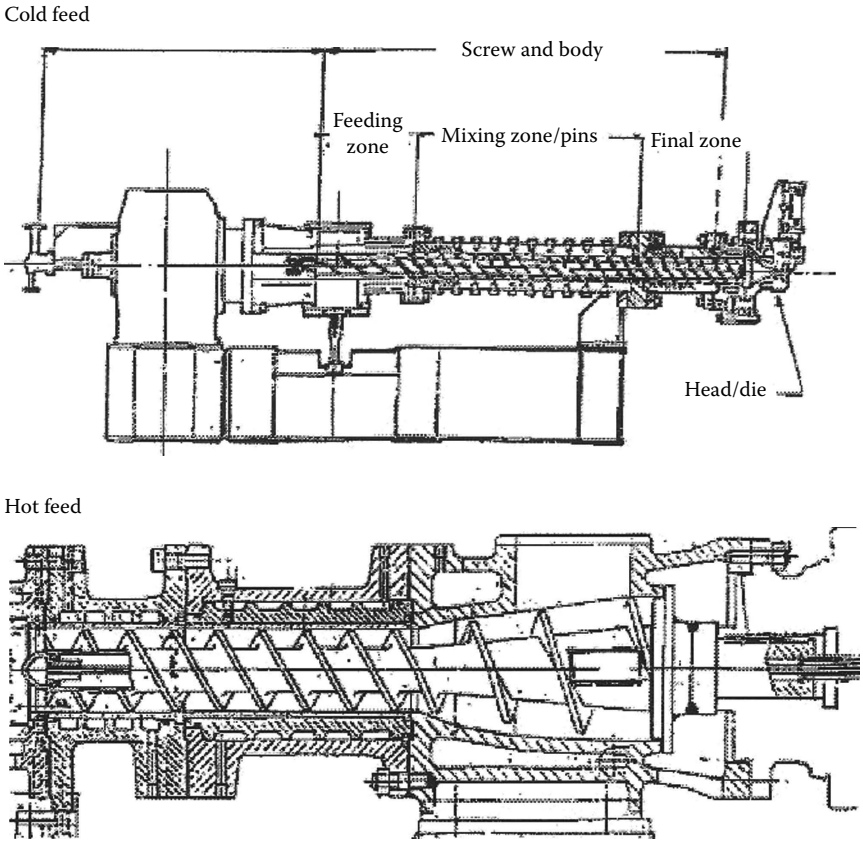


FIGURE 19.10 Complete scheme of hot feed and cold feed extruders. Cold feed: length of the screw from 12 to 16 times its diameter. The screw is divided into three zones: (1) feeding/supply zone, (2) mixing zone due to central pins, and (3) anti-pulsation end zone. Hot feed: length of the screw from 6 to 8 times its diameter. The screw is divided into two zones: (1) conic feeding zone and (2) cylindric extrusion zone.

19.3.3 Building

As the semifinished materials, such as liner, plies, beads, belts, sidewalls/rim cushion, and tread, have been obtained by the previous processes, the next step is the tire building. It can be divided into two phases: first stage and second stage.

19.3.3.1 First Stage

This phase of the manufacture consists of building up the tire by laying down on the surface of a cylinder the different parts of the tire structure (see Section 19.4): liner, carcass plies, beads and relative fillers, and sidewalls and rim cushions. To avoid the presence of air between the parts, a roller process is performed: by means of pressure, the air is forced to flow out to form a homogeneous set.

The obtained manufacture looks like a sleeve, called *first phase carcass* (see Figure 19.11a).

19.3.3.2 Second Stage (Conformation)

The output of the first stage is a product that has a cylindrical shape. The second stage consists of laying down on a drum (with higher diameter than the first stage) the two steel belts, nylon 0° and the tread band (Figure 19.16). By means of conformation, the final shape of the tire (close to a toroid) is obtained. The beads (Section 19.4.3) are moved one toward the other (Figure 19.11b) by means of inflation.

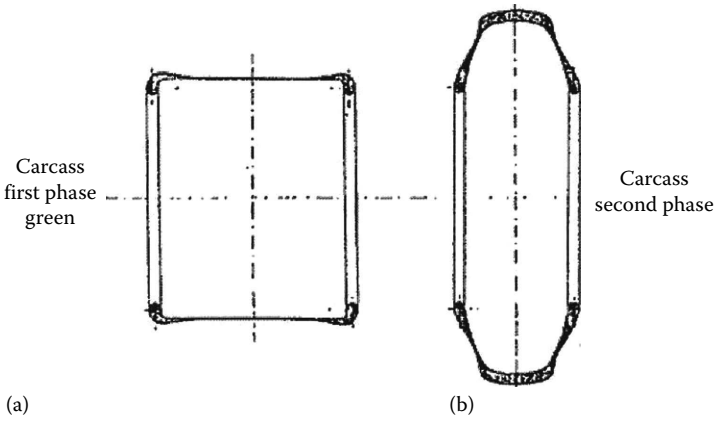


FIGURE 19.11 Phases of carcass creation. (a) Confection. (b) Conformation.

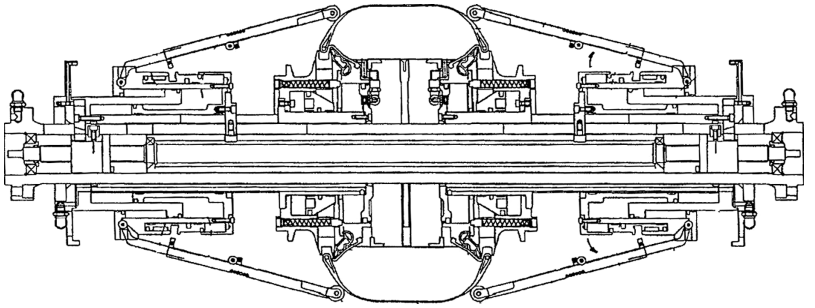


FIGURE 19.12 Confection machine. In the middle the tire can be seen.

Like in the first phase, after conformation, there will be a rolling operation around this new shape. At the end of this step, the tire is called *uncured tire* or *green tire*, ready to be cured to obtain the final product.

Figure 19.12 shows a second stage tire-building machine during the conformation phase.

19.3.4 Curing

19.3.4.1 Introduction to Curing

Curing is the process of vulcanization. The unvulcanized rubber can be considered a set of polymer chains disposed randomly in the space (Figure 19.1). If submitted to deformation, they will not return to their original shape. Vice versa, the vulcanized rubber [15–17] will return to the original shape after the deformation is ended, due to cross-links created by sulfur (Figure 19.1).

The main problem related to the curing process is how to determine the optimum amount (or density) of cross-links. Such a density is closely related to the heat provided during curing.

The final properties of the rubber compound depend on cross-link density and type of links (Figure 19.13). Such properties are hysteresis, tear strength, tensile strength, and stiffness. They are defined, respectively, as follows (see Figure 19.13).

Hysteresis is related to the ability of a material to come back to its original shape when the load is removed. The more the cross-links, the less the hysteresis. The higher the hysteresis, the higher the grip of the tire on the ground.

Stiffness is directly related to the modulus of elasticity (Young modulus). The more the cross-links, the more the stiffness.

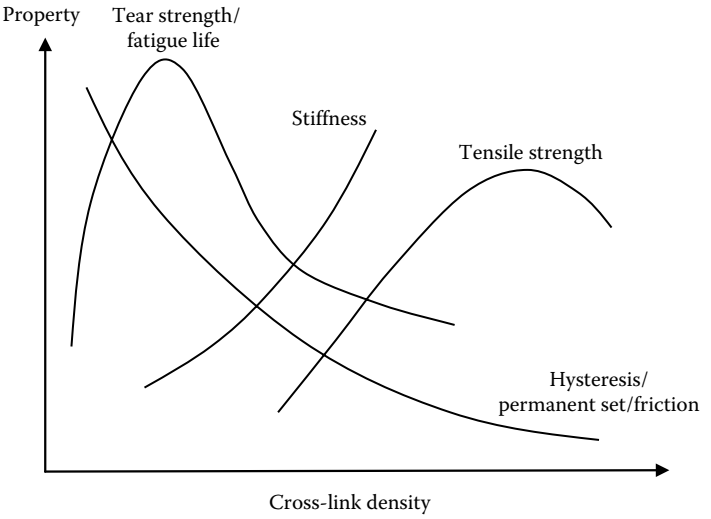


FIGURE 19.13 Effects of cross-links obtained in the rubber by vulcanization process.

Tensile strength reaches a maximum at a relatively high cross-link density. This is due to the brittle nature of a rubber with an excessive cross-link density. Actually in this case, the nature of the cross-link changes.

Tear strength reaches a maximum at a relatively low cross-link density.

In the curing phase, the external part of the green tire is subjected to high temperature and high pressure: it is pressed against the mold so that the tread pattern is created.

In the internal part of the green tire, the cohesion of all the components is also achieved.

19.3.4.2 Curing Cycle

Ideally, constant values of temperature and pressure during the curing process could be prescribed. However, for actual applications, the vulcanization is based on time-varying temperature and pressure. The final vulcanization result (cross-link density) depends on the energy that is provided. However, the nature (mono-, di-, or polyatomic chains) of the sulfur bridges depends strongly on the curing cycle (pressure–temperature and time histories). A variation of 10° can double or halve the curing time, producing different sulfur bridges.

19.3.4.3 Curing Mold

A tire-manufacturing mold has got a pattern corresponding to the outer profile of the tire being molded. The pattern influences steering, wear, roadholding, and noise characteristics of the tire.

The curing apparatus consists of many components, such as bladder, steam supply, and mold (Figures 19.14 and 19.15). The outer surface of the raw tire is closely in contact with the face of the mold; this is due to the action of the bladder acting on the inner surface of the tire. After the curing cycle, the tire pattern (and the whole tire) will have the desired shape.

According to the production volume, basically two materials are used for the mold, aluminum or steel: The former, for low amount of production, facilitates forming the deep and complex features, and the latter has got higher strength and durability.

19.3.4.4 Lubricant for Curing Cycle

A relative motion occurs between the outer contact surface of the bladder and the inner surface of the tire. Such a motion happens during the expansion phase (bladder inflated) and at the end of the curing cycle when the bladder has to come back to the original configuration (bladder deflated). Lubrication between the surfaces experiencing the relative motion is essential to avoid buckle

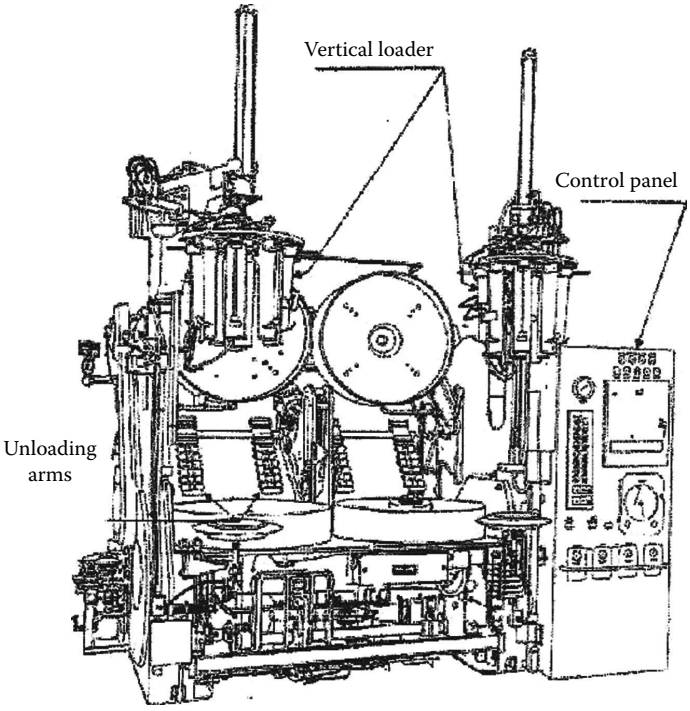


FIGURE 19.14 Curing machine. The green tire is positioned inside the molds (the four cylindrical parts visible) and then pressure and temperature are raised up to the desired values.

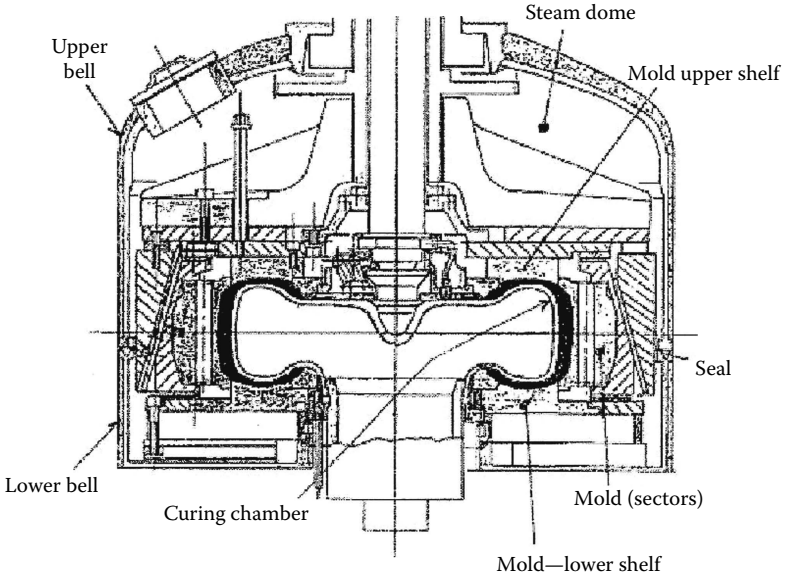


FIGURE 19.15 Curing chamber. The upper and lower parts of the mold (upper shelf and lower shelf) are joined to contain the green tire.

(a phenomenon that refers both to defects of the shape of the tire and wear problem on the bladder). If lubrication, heat, and pressure are not sufficient during the curing process, air may be trapped inside the tire structure.

The green tire is pre-lubricated to avoid buckle. The majority of lubricants are water based and can contain elements such as bentonite clay, polydimethysiloxane, sodium benzoate, mica, and kaolin clay.

19.3.4.5 Spews

When the green tire is pressed versus the mold, there is air between them. Since the pattern is made with protrusion from the surface of the mold, the air must have the possibility to escape. If air is trapped, on the surface of the cured tire, small dimples can appear, known as shortage, but also bubbles may be formed. To discharge the air between the raw tire and the mold, small vent holes, with usual diameter from 0.1 up to 2 mm, are created in the mold faces; such vent holes cause spews on the surface of the vulcanized tire. Spews are the unwanted small projections on the tire outer surface.

Spews formed on the tread face will be easily removed during running. On the contrary, spews on the sidewall will not be removed.

The mold can be composed by sectors, and between neighborhood surfaces micro-clearances remains. From these clearances air can escape, while the rubber cannot flow inside. Since they are extremely thin, bubbles would not be produced on the surface of the tire.

19.4 TIRE STRUCTURE

When a tire has to be designed, the aim is to satisfy a so-called *target letter* coming from a car manufacturer. Alternatively, generic specifications for a tire can be derived referring to the requirements from the market. Design targets are physical and mechanical performances, in particular mileage, rolling resistance, noise emission, wet grip, and others. Based on such specifications, the tire is implemented and tested (Figure 19.16).

In the past, tire companies worked without a standard on the final product. However, to guarantee a constant quality product, the companies issued (and issue) inside standards for defining the desired production cycle to be followed. Such internal standards created a basis for quality control

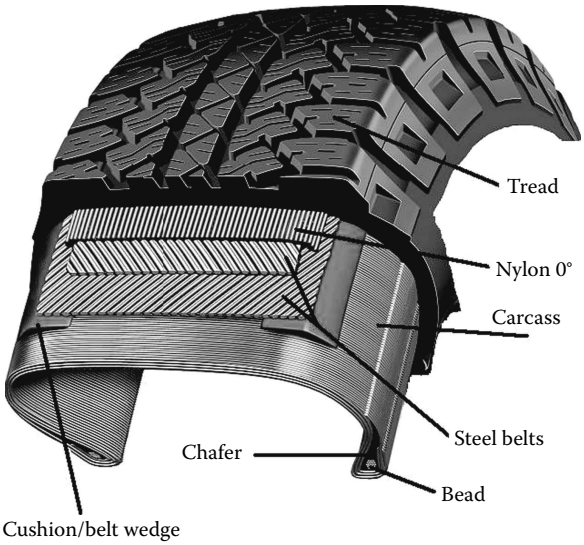


FIGURE 19.16 Tire nomenclature. (Adapted from http://www.avontyres.com.au/technical/images/cutaway-passenger.jpg)

on the whole manufacturing process. Quality control led to a more efficient production and to more tight product tolerances.

The first standards constrained the dimension of the tire. The inflated tire has to be bounded in both radial and transversal directions. TRA [18] and ETRTO [19] issued important standards on tire dimensions.

DOT/ECE (see [7] for an overview of the topic) issued important standards on tire safety.

The addressed standards are aimed to let companies adopt uniform technical prescriptions for wheeled vehicles; competition tires are an exception.

Relatively recently [20,21], a European legislation was introduced to guarantee the quality of the final product.

19.4.1 Inner Liner

The inner liner (Figure 19.17) is usually made with a compound based on *halobutyl* rubber; this polymer cures faster compared with *butyl*, and it needs smaller amount of curative agents. It can also be co-cured more readily in contact with other elastomer making a good rubber product. The inner liner prevents gases from migration into the carcass plies or towards the belt plies of the tire (Figure 19.18). It looks soft and has a relatively low tear strength and elastic modulus compared with the other rubbers used in the tire. The inner liner has to be limited in nonstressed areas of the inner tire. Inner liner should have a good moisture permeability, a good adhesion to carcass compounds, flex fatigue resistance, and durability. The air barrier properties and flex fatigue result from low level of cross-link density.

19.4.2 Plies

Plies are made from rubber compound and textile cords (Section 19.2.7.2) in order to give the needed structural reinforcement (Figure 19.19). Plies encircle the bead (Figure 19.19) and they are

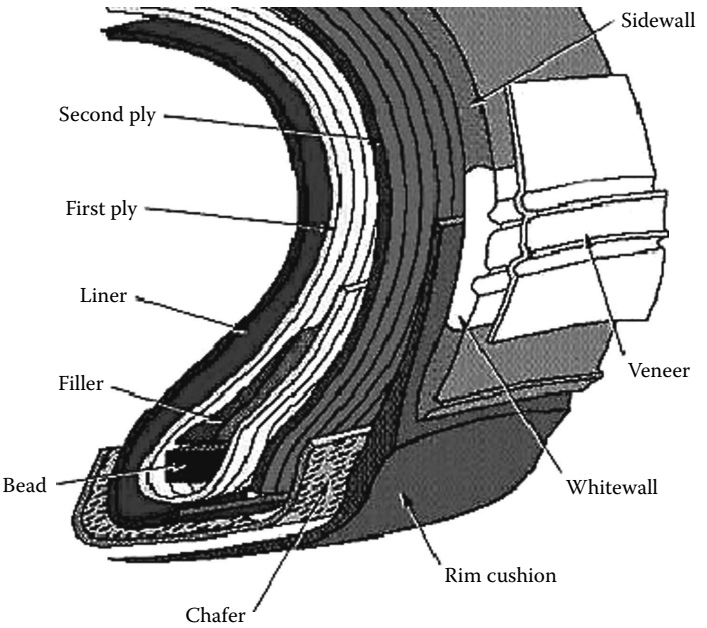


FIGURE 19.17 Tire nomenclature for inner parts. (Adapted from http://www.2405.com/press-library/Hummer-H3-2006/cutaway-of-the-bridgestone- lt285-75r16c-33-inch-on-off-road-radial-tire.jpg)

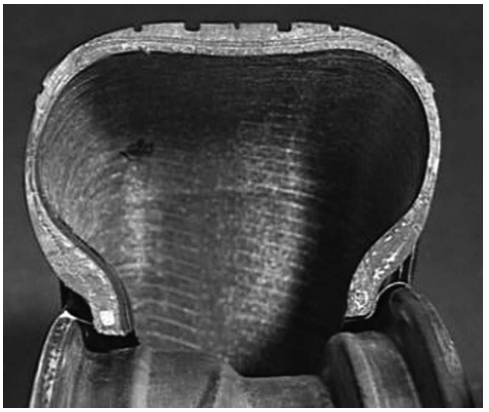


FIGURE 19.18 Tire section.

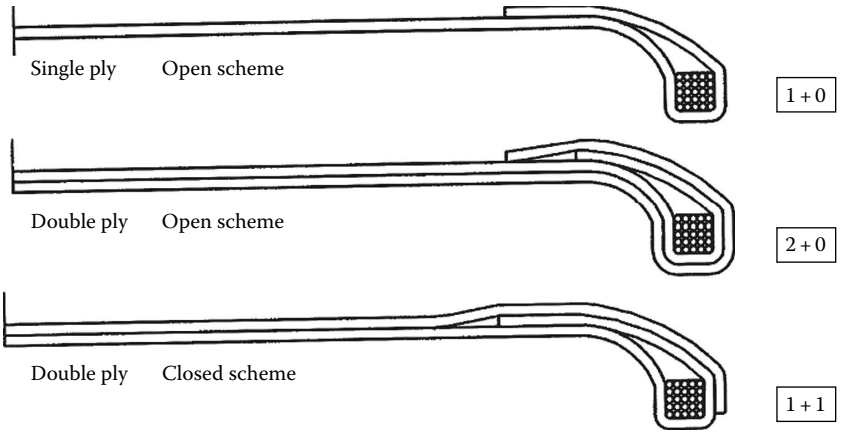


FIGURE 19.19 Schemes of plies forming the bead.

positioned by special machinery. A ply is a highly flexible part, but due to the internal pressure of the tire it can support the weight of the car (see Chapter 11). Plies embody the carcass of the tire.

19.4.3 Bead and Filler

The beads are made from high-resistance steel, according to ASTM 1070. Their function is to fit the pneumatic tire on the rim (Figure 19.20).

The bead filler is the most stiff compound in the tire. It constraints the bead wire to maintain its desired position. Due to this, the beads do not affect wheel balance, as they are firmly constrained to the rim.

19.4.4 Chafer and Rim Cushion

The chafer consists of a thin layer of fabric placed around the beads (Figure 19.17). The aim is to protect the bead steel cords against unexpected rise of problems such as wear and cutting action due to the contact with the rim. Due to its location, it can also distribute the stresses. Additionally, it is a boundary from external undesired particle of dirt or moisture that can flow inside the tire. Sometimes it is absent.

The rim cushion is made from a compound with high grip in order to avoid slipping on the rim. It provides the necessary air tightness to prevent leakage.

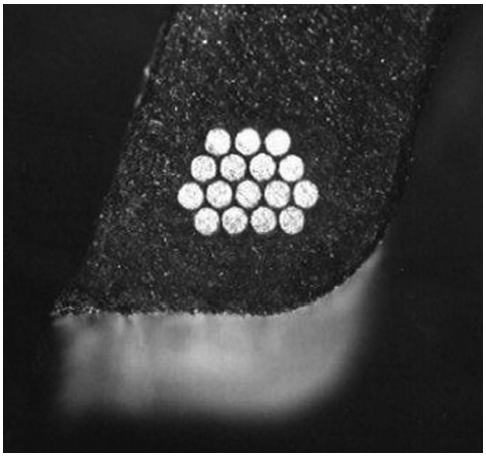


FIGURE 19.20 Bead cross section.

19.4.5 Sidewall

It is placed in between the beads and the tread, in the side outer part (Figures 19.16 and 19.17). It is made usually from natural rubber (NR) and polybutadiene (BR). It covers the carcass and protects it from abrasion. It has good fatigue resistance, low hysteresis, and low modulus of elasticity.

19.4.6 Belts

Belts are placed under the tread over the carcass of the tire (Figure 19.16), and they can be more than one and are made from steel, coated with rubber. The functions of such components are various. They strengthen the whole tire structure constraining the carcass cords to keep their radial desired position. They resist to the hoop tension (the prestress directed tangentially with respect to the tread, due to tire inner pressure). They also resist to a puncture from an external object.

The steel cords are sloped with respect to the meridian plane of the tire by an angle ranging from 20° to 35°. This is due to resist side forces. Usually, passenger tire has two belts whose slopes are opposite. Such an anisotropy is directly related to ply steer [3]. To avoid the buckling of the belt package, the cord size and the cord encount (number of cords per cross width) are properly chosen.

19.4.6.1 Nylon 0°

The nylon belt is placed between the belt package and the tread and improves tire safety. Primarily it resists centrifugal forces. It improves tire uniformity and isotropy. It resists the hoop tension [3]. Experience states that a better ride comfort is provided by tires with nylon 0° belt (Figure 19.21).

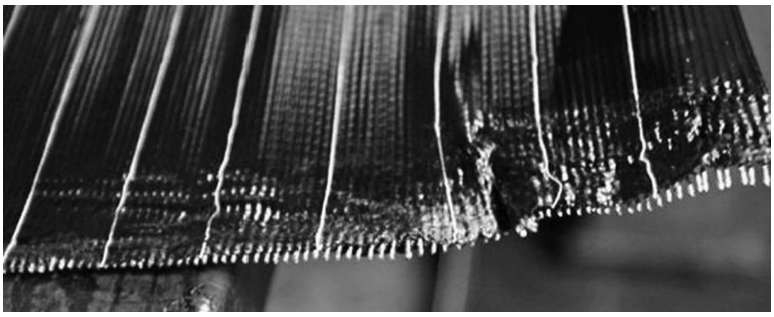


FIGURE 19.21 Nylon belt.

19.4.7 Belt Wedge: Edge Cover

This component is not present in every tire, sometimes it is needed for safety and durability improvement. It is placed at both sides of the belts (Figure 19.22). Joining together the edge parts of the belts, it strengthens them and so improves tire fatigue resistance. It also prevents dangerous phenomena like belt separation.

19.4.8 Tread

The tread is the outer part of the tire in contact with the ground. It is placed upon the belts (Figure 19.23). The designer, according to the mission of the tire, defines the proper rubber compound for the desired wear resistance. Also, other performances are obtained mainly by the proper choice of rubber compound.

The design of ribs and grooves is of primary importance for tire dynamics. To avoid aquaplaning, the grooves must be either longitudinal or sloped. Traction is improved if groves in lateral position are transversal. The noise perceived inside the vehicle is related to the spacing of the transversal grooves. An unequal spacing flattens and enlarges the noise spectrum. Moreover, the correct phase of the two lateral sets of groves along the shoulders reduces the amplitude of the spectrum.

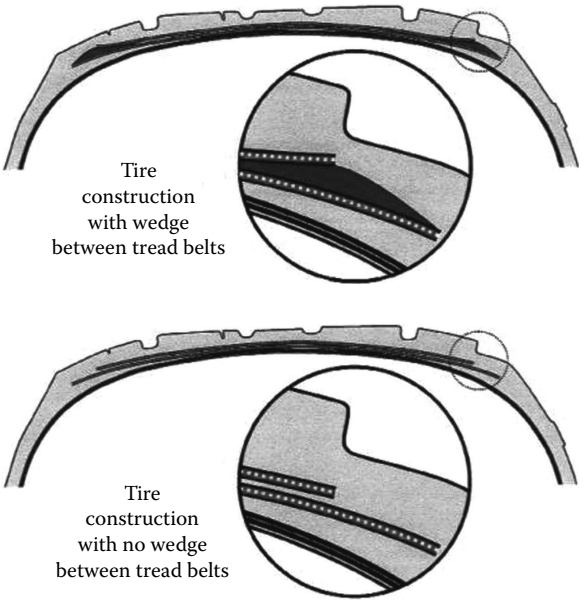


FIGURE 19.22 Wedges between belts. (Adapted from www.tirelitigation.com)



(a) (b)

FIGURE 19.23 Treaded (a) and slick (b) tires.

The variety of requirements for different car segments drives the design of the tread, starting from off-road tire up to slick. Even in the slick tire, the tread is present.

Road car tires need a compromise between mileage and dynamic performances. Strongly varying temperature conditions (hot summer, cold winter), water on the road, and snow have to be taken into account. The ratio between the area of voids and the rubber in the tread (ratio "land and sea") is approximately ranging between 25% and 30% for passenger summer tires. For winter tires (M + S), the ratio is 35%–40%.

Another important aspect is ecology. The tire tread greatly influences (together with the compound and the geometrical dimensions—i.e., curvatures) the rolling resistance. Typical values are 10 N in longitudinal direction per 1000 N in radial direction, the minimum being 7 N and the maximum 12 N.

Tread wear indicators are placed according with the UN/ECE standard [22,23]. The purpose is to give an idea of the residual tread when its height is less than 1.6 mm.

19.5 TIRE STANDARDS AND TESTS

19.5.1 UN/ECE Regulations

The UN/ECE 1958 Agreement [22] on "vehicle type approval and its mutual recognition" is an international forum (promoted by the United Nations) providing the legal framework under which tire technical regulations are developed. More than 50 countries from various regions of the world including the European Community have adhered to the UN/ECE 1958 Agreement.

The UN/ECE 1998 Global Agreement purpose is to further enhance the process of international harmonization through the development of global technical regulations (GTR), which may cover countries that are not contracting parties to the 1958 Agreement.

The European tire industry initiated the development of a GTR for tires in close cooperation with Japanese and U.S. colleagues [10].

The main "tire" technical regulations (useful for product development and certification) are [10]

- UN/ECE Regulation 30: tires for motor vehicles and their trailers [22,23]
- UN/ECE Regulation 54: tires for commercial vehicles and their trailers
- UN/ECE Regulation 64: temporary-use spare wheels/tires
- UN/ECE Regulation 75: tires for motor cycles and mopeds
- UN/ECE Regulation 106: tires for agricultural vehicles and their trailers
- UN/ECE Regulation 117: tires rolling sound emissions and adhesion on wet surfaces
- UN/ECE Regulation 108: retreaded tires for motor vehicles and their trailers
- UN/ECE Regulation 109: retreaded tires for commercial vehicles

ECE standards introduced the so-called tire type, a set of categories in which the tires can be classed according to

- The manufacturer
- Tire size
- Category of use (ordinary [road-type] or snow tire or for temporary use)
- Structure (diagonal [bias-ply], bias-belted, radial-ply, run-flat tire)
- Speed category
- Load capacity
- Tire cross section

In Table 19.3, the possible structural characteristics for tires are reported.

TABLE 19.3
UN/ECE 30 Table 1

Structure	Angle of Ply Cord	Belt under Tread	Extra Reinforced Carcass	Temporary-Use Notation	Flat Tire Running Mode
Diagonal or bias-ply	Less than 90°	X	X	X	X
Bias belted	Less than 90°	V	X	X	X
Radial	90°	V	X	X	X
Reinforced or extra load	X	V	V	X	X
Temporary-use/spare tire	X	Not applicable	X	V	X
T-type temporary-use spare tire	X	Not applicable	X	V Notice: high pressure	X
Run-flat tire	X	V	V	X	V

Source: United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, 2007.
X = no, V = yes.

19.5.1.1 Markings

Each tire (except motorsport tires and other tires for special use) has to show on the sidewall specific symbols grouped into four categories (A, B, C, D) as shown in Figure 19.24.

The rim can be classified as theoretical rim, design rim on which the final tire is supposed to be mounted on, or test rim, the ones used for testing; see for further details [19].

In Tables 19.4 through 19.6, the lists of load capacity and speed indices are, respectively, reported.

If the speed is greater than 210 km/h, the load index (Li) is modified according to the values reported in Table 19.6. For intermediate speeds, a linear interpolation is accepted.

If the speed is lower than 60 km/h, the load index (Li) is modified according to the values reported in Table 19.7.

The values reported in the preceding tables guarantee the basic tire functions, that is, carrying a given load at specific speed, steering, and applying the traction/braking forces to the ground.

19.5.1.2 Tire Dimensions

In the ECE standard Regulation 30 [22,23], the tire dimensions are defined. The section width (Figure 19.25) depends on the rim on which the tire is fitted. The relationship between the section width and the rim width is given by the formula

$$S = S_1 + k(A - A_1)$$

where

S is the actual "section width" expressed in mm
S_1 is the "nominal section width" (in mm) as shown on the sidewall of the tire in the designation of the tire as prescribed
A is the width (expressed in mm) of the actual rim
A_1 is the width (expressed in mm) of the reference rim
k shall be taken equal to 0.4, more details on k are reported in the standard [23]

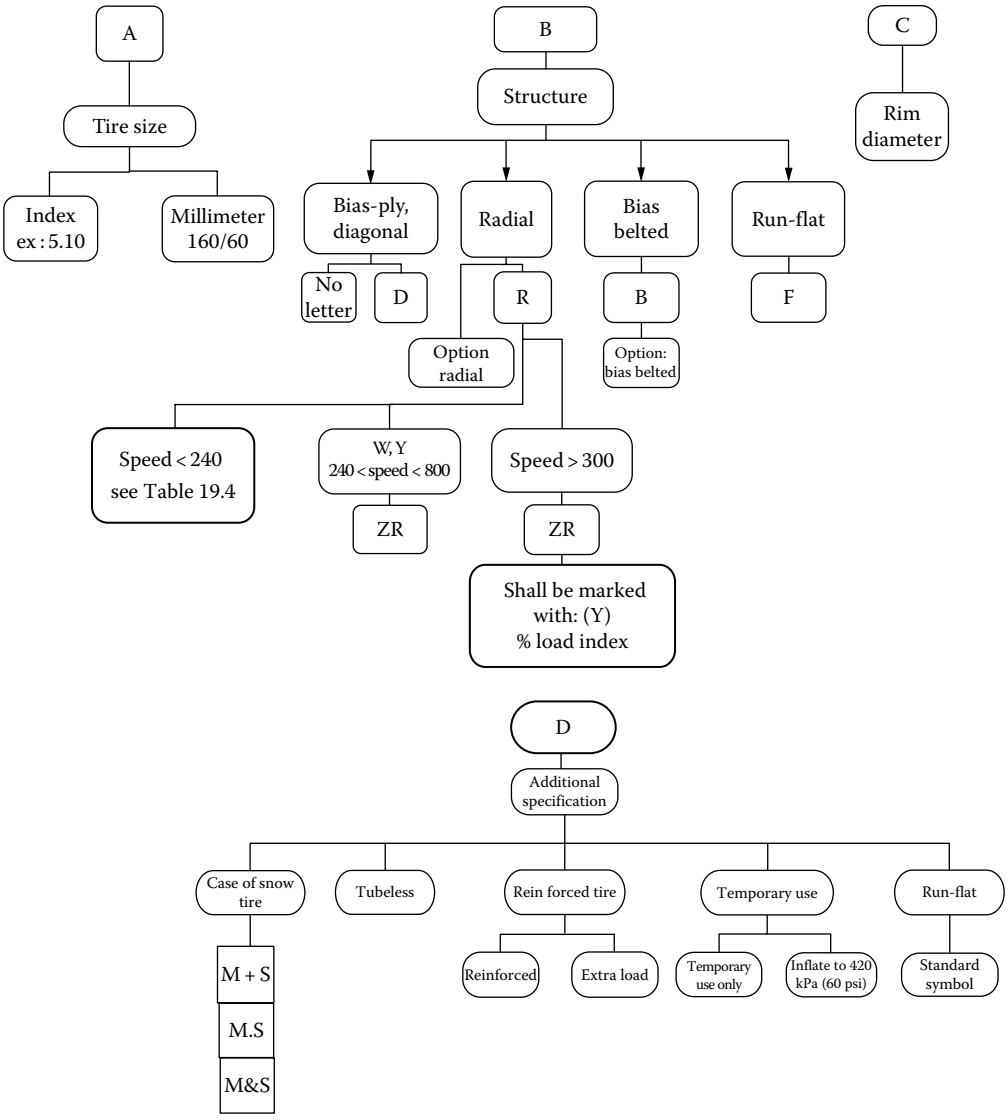


FIGURE 19.24 UN/ECE tire marking scheme.

TABLE 19.4
ECE Speed Category Symbol

Speed Category Symbol	Maximum Speed (km/h) (mph)	Speed Category Symbol	Maximum Speed (km/h) (mph)	Speed Category Symbol	Maximum Speed (km/h) (mph)
L	120 (75)	R	170 (106)	H	210 (130)
M	130 (81)	S	180 (112)	V	240 (149)
N	140 (87)	T	190 (118)	W	270 (168)
P	150 (93)	U	200 (124)	Y	300 (186)
Q	160 (99)				

Source: United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, 2007.

TABLE 19.5
Load Capacity Indices

Li	kg	Li	kg	Li	kg	Li	kg
0	45	31	109	61	257	91	615
1	46.2	32	112	62	265	92	630
2	47.5	33	115	63	272	93	650
3	48.7	34	118	64	280	94	670
4	50	35	121	65	290	95	690
5	51.5	36	125	66	300	96	710
6	53	37	128	67	307	97	730
7	54.5	38	132	68	315	98	750
8	56	39	136	69	325	99	775
9	58	40	140	70	335	100	800
10	60	41	145	71	345	101	825
11	61.5	42	150	72	355	102	850
12	63	43	155	73	365	103	875
13	65	44	160	74	375	104	900
14	67	45	165	75	387	105	925
15	69	46	170	76	400	106	950
16	71	47	175	77	412	107	975
17	73	48	180	78	425	108	1000
18	75	49	185	79	437	109	1030
19	77.5	50	190	80	450	110	1060
20	80	51	195	81	462	111	1090
21	82.5	52	200	82	475	112	1120
22	85	53	206	83	487	113	1150
23	87.5	54	212	84	500	114	1180
24	90	55	218	85	515	115	1215
25	92.5	56	224	86	530	116	1250
26	95	57	230	87	545	117	1285
27	97.5	58	236	88	560	118	1320
28	100	59	243	89	580	119	1360
29	103	60	250	90	600	120	1400
30	106						

Li, load capacity index; kg, mass of the vehicle resting on the single tire that is to be carried.

The outer diameter can be calculated with the formula

$$D = d + 2H$$

where

D is the outer (overall) diameter expressed in mm
d is the conventional rim diameter expressed in mm
H is the nominal section height in mm, equal to $H = 0.01S_1 \cdot R_a$, where S_1 is the nominal section width in mm and R_a is the nominal aspect ratio, as shown on the sidewall of the tire in the tire size designation

Higher values of the width of the manufactured tire are accepted up to 6% or 4%, respectively, for bias-ply or radial/run-flat tires. If the tire has specific ribs for rim protection, the width can exceed 8 mm than the nominal value.

TABLE 19.6
Speed Category Symbol for Speed Greater than 210 km/h

Speed Category Symbol	Max Speed	Max Load Rating (%)
V	215	98.5
	220	97
	225	95.5
	230	94
	235	92.5
	240	91
W	240	100
	250	95
	260	90
	270	85
Y	270	100
	280	95
	290	90
	300	85

TABLE 19.7
Maximum Load for Speed Lower than 60 km/h

Max Speed (km/h)	Max Load Rating (%)
25	142
30	135
40	125
50	115
60	110

The external (outer/overall) diameter of the tire must not be outside the values D_{min} and D_{max} obtained from the following formulae:

$$D_{min} = d + 2Ha$$

$$D_{max} = d + 2Hb$$

where

H and d have been defined earlier (Figure 19.25), otherwise

$H = 0.5 \cdot (D - d)$

$a = 0.97$

$b = 1.04$ (radial or run-flat tires) or $b = 1.08$ (diagonal or bias-belted tires)

For snow tires, the maximum allowable outer diameter is $1.01D_{max}$.

In the Annex 6 of UN/ECE Regulation 30 [23], a procedure is given to measure the dimensions of a tire.

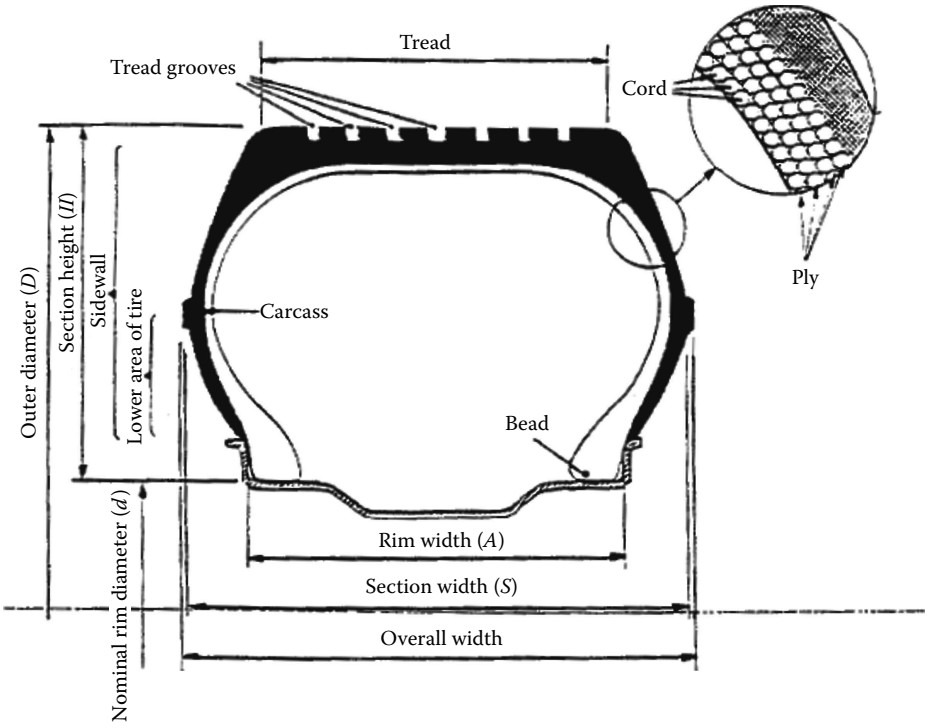


FIGURE 19.25 Tire dimensions. (From United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, 2007.)

19.5.1.3 Load and Speed Performance Test

In the Annex 7 of UN/ECE Regulation 30 [23], a procedure is given to test the integrity of the tire.

The tire has to be mounted on the test rim, then inflated to pressure according to Table 19.8. For T tires (temporary use), a pressure equal to 4.2 bar is adopted.

The procedure for assessing the tire integrity is reported in Figure 19.26. Basically, the tire is inflated, loaded at 80% of the load index. Then the speed is rumped up near the maximum speed

TABLE 19.8
Inflating Pressures for Load and Speed Performance Test

	Diagonal (Bias-Ply) Tires			Radial/Run-Flat System		Bias Belted Tires
	Ply Rating					
Speed Category	4	6	8	Standard	Reinforced	Standard
L, M, N	2.3	2.7	3.0	2.4	2.8	—
P, Q, R, S	2.6	3.0	3.3	2.6	3.0	2.6
T, U, H	2.8	3.2	3.5	2.8	3.2	2.8
V	3.0	3.4	3.7	3.0	3.4	—
W	—	—	—	3.2	3.6	—
Y	—	—	—	3.2	3.6	—

Source: United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, Annex 7, 2007.

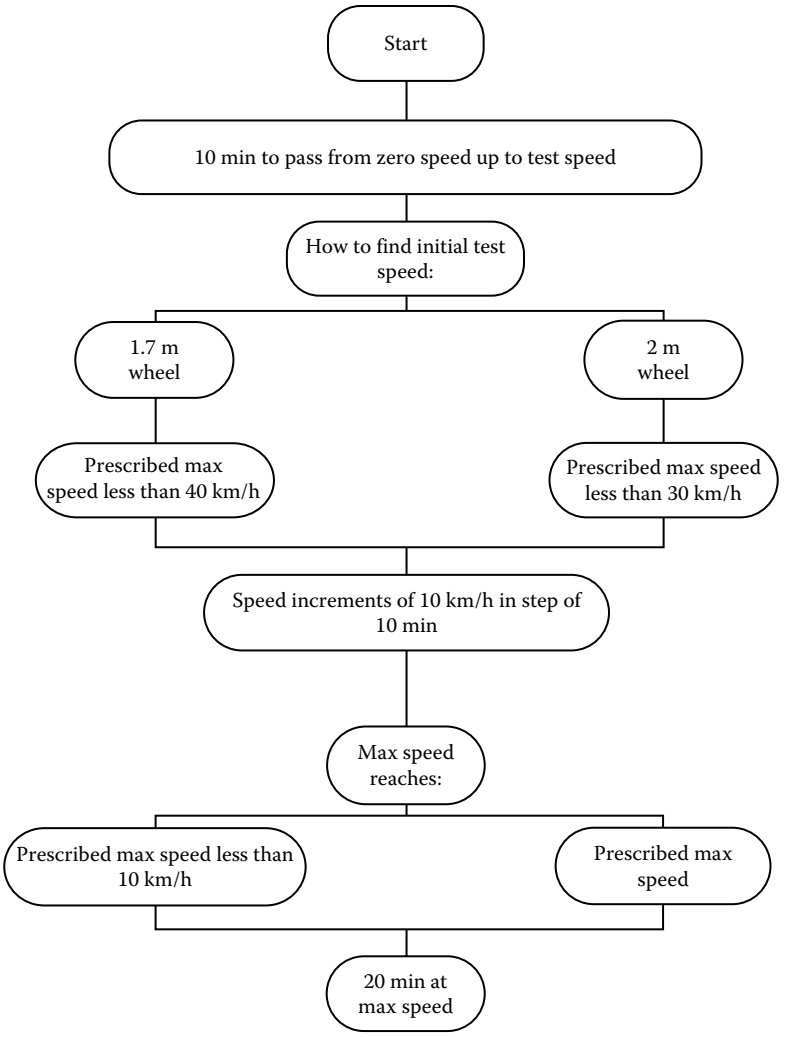


FIGURE 19.26 Load and speed performance test as defined in UN/ECE Regulation 30, E/ECE324. (From United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, 2007.)

and then increased by steps of 10 km/h (lasting 10 min) up to the very maximum speed, defined by the speed index (such speed has to be maintained for 20 min). Exceptions (and technicalities) are reported in [23] and in Figure 19.26. For example, a test performed with a wheel (drum) of 1.7 m with a radial tire marked H (speed index_210 km/h) and load index 100 (800 kg) is organized as follows: the tire is inflated at 2.8 bar, loaded at 640 kg; in 10 min the speed is raised at 170 km/h; and then the speed is increased at 180 (for 10 min), 190 (for 10 min), and 200 (for 20 min).

For tires belonging to the Y speed category (see Figure 19.24), a special test is performed to assess integrity at high speed (see [23], Annex 7).

A tire which after undergoing the load and speed test does not exhibit any tread separation, ply separation, cord separation, chunking, or broken cords shall be deemed to have passed the test. Chunking means the breaking away of pieces of rubber from the tread, cord separation means the parting of the cords from their rubber coating, ply separation means the parting of adjacent plies, and tread separation means the pulling away of the tread from the carcass.

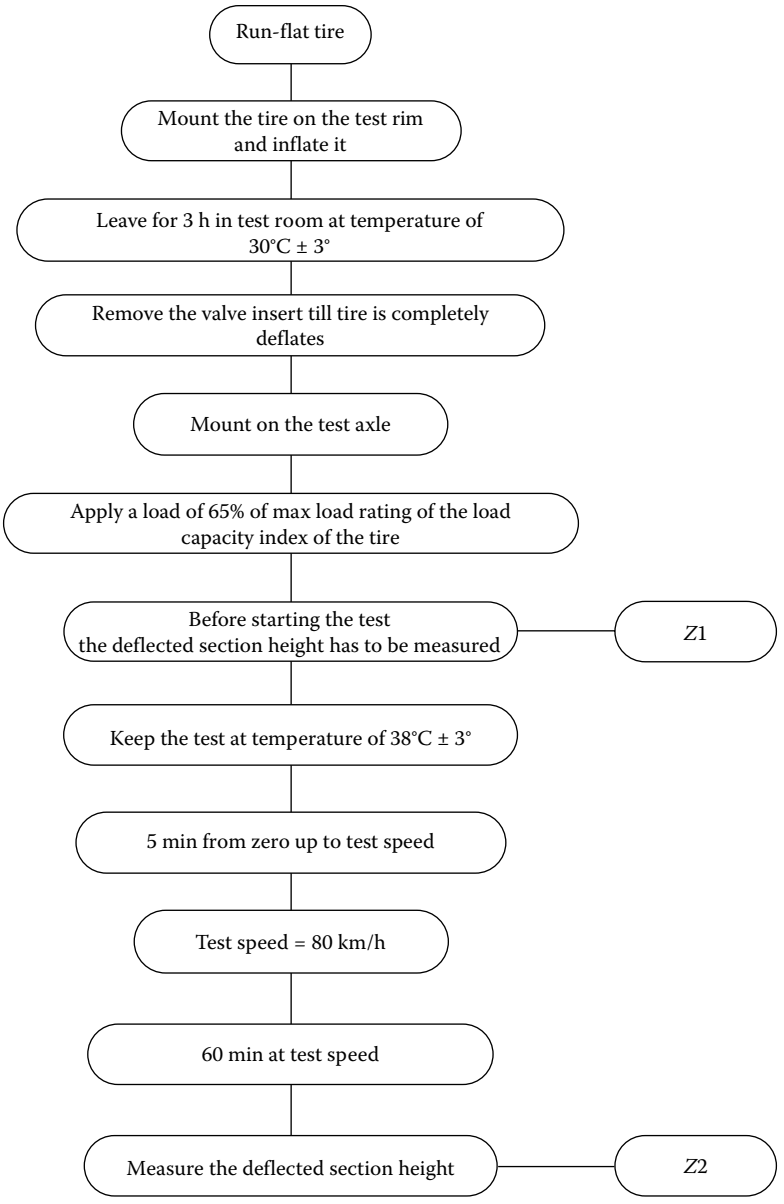


FIGURE 19.27 Integrity test for run-flat tires. (From United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, Annex 7, 2007.)

The outer diameter of the tire, measured 6 h after the load and speed performance test, must not differ by more than ±3.5% from the outer diameter as measured before the test.

Y speed category tires pass the test even if they show blistering on the tire tread. For run-flat tires, in addition to the test described in Figure 19.27, the integrity is tested as shown in Figure 19.27, which refers to the deflated running condition. At the end of the test, the measure of the deflected section height ($Z2$) is taken. The change in percent of the deflected section height compared to the deflected section height at the start of the test ($Z1$) is calculated: $([Z1 - Z2]/Z1) \times 100$ (the deflected section height is measured according with the standard in [24]). If such a figure is higher than 20% and no other defects are visible, the tire is deemed to have passed the test.

19.5.1.4 Approval of the Tire

The approval corresponds to the homologation of the tire. The public administration certifies that the manufacturer has obtained a product that complies with the standards. The historical certifications are obtained from DOT or ECE.

19.5.2 DOT Tests FMVSS139

DOT [25,26] has issued integrity tests similar to the ones reported in the previous section [27]. Unseating resistance, strength, endurance, and high-speed performance tests are prescribed. Three specimens have to be used for one tire, one to test bead unseating and strength, one for tire endurance, and the last one for high-speed test.

19.5.2.1 Tire Dimensions

The dimensions of the tire have to be bounded, in accordance with the inflating pressure specified by the manufacturer, according to Figure 19.28.

19.5.2.2 Bead Unseating Test

The tire has to be clean, dry in the beads, and mounted at the specified pressure on a painted and clean rim, without use lubricant or adhesives. For this test a specific structure is provided, as shown in Figure 19.29.

The test has to be carried out applying a load at a specified distance from the rim on the sidewall, the rate of increase is usually set to 2 in./min (i.e., 50 mm/min), and the initial position of the load arm is parallel to the tire and rim. The test stops if the beads unseating or the value specified for the load is reached. The procedure has to be repeated in order to cover a 90° shift between each test.

The test is performed at a pressure lower than the reference one. The load is applied at a defined position on the sidewall. The loads are reported in [27] and, depending on tire size, are bounded in the range 6,700–11,100 N.

19.5.2.3 Tire Strength

The test consists of forcing a 3/4 in. (19 mm) diameter cylindrical steel plunger with hemispherical end perpendicularly into the tread rib, as close to the center of the tire as possible, avoiding penetration into tread groove, with a rate of 2 in./min. The test is repeated for every 72°. The tire is inflated at the pressure adjusted as specified in [27].

The force is recorded in kilograms (lb), penetration is recorded in millimeters (inches), and energy is computed in joules (inch pounds) for each load application. If the tire fails to break before

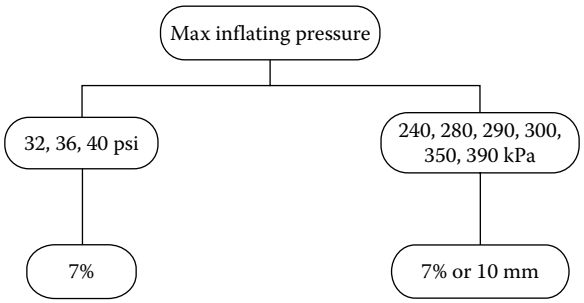
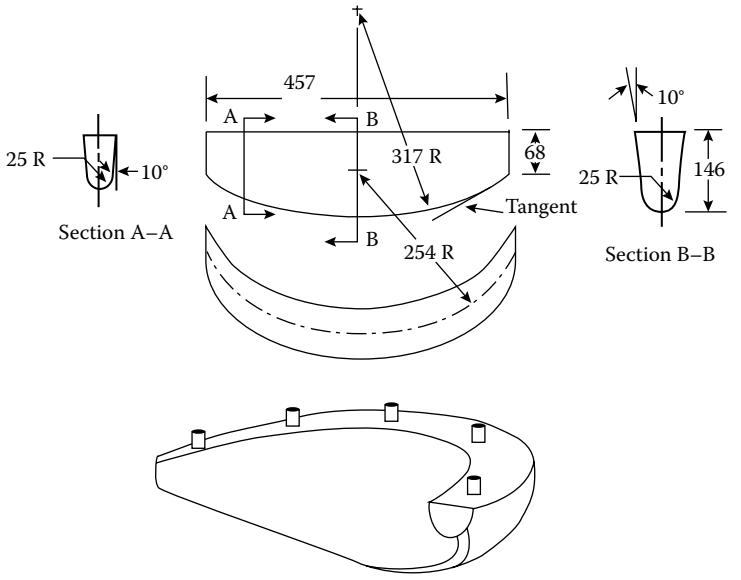


FIGURE 19.28 Max inflating pressure DOT tests.



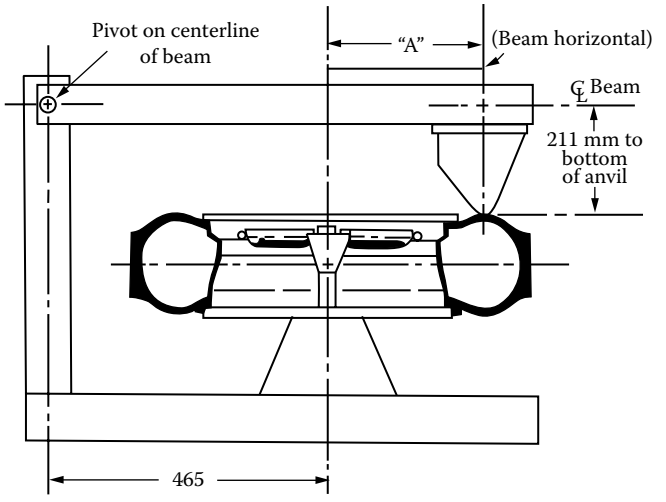


FIGURE 19.29 Bead unseating test device.

the plunger is stopped by reaching the rim, record the force and penetration as the rim is reached and use these values in the formula indicated in the following:

$$W = \frac{FP}{2}$$

where

W is the energy (inch pounds)
F is the maximum force (pounds)
P is the maximum penetration (inches)

The final value of the breaking energy is the average between the five tests.

TABLE 19.9
Requirements for Tire Strength Test (Minimum Static Breaking Energy)

Load Range / Time Characteristic	Light Truck, 17.5 in. or Smaller Rim Tubeless		Tube Type		Tubeless	
Plunger Diameter (mm and in.)	19.05 J	¾* in. lb	31.75 J	1-1/4* in. lb	J	in. lb
A	225	2,000				
B	293	2,600				
C	361	3,200	768	6,800	576	5,100
D	514	4,550	892	7,900	734	6,500
E	576	5,100	1,142	12,500	971	8,500
F	644	5,700	1,785	15,800	1,412	12,500
G	711	6,300				
H	768	6,800				

Load Range / Time Characteristic	Tube Type		Tubeless	
Plunger Diameter (mm and in.)	(38.10) J	(11/2*) in. lb	J	in. lb
G	2,282	20,200	1,694	15,000
H	2,598	23,000	2,090	18,500
J	2,824	25,000	2,203	19,500
L	3,050	27,000		
M	3,220	28,500		
N	3,389	30,000		

Source: NHTSA, Laboratory test procedure for FMVSS no. 139, New pneumatic radial tires for light vehicles, Washington, DC, January 2006.

Note: For rayon cord tires, applicable energy values are 60% of those in table.

If none of the first four applications has resulted in carcass breakage below the minimum energy value, the fifth (last) plunger application is to be taken beyond the minimum limit to casing failure or to the point of tire compression against the rim, whichever occurs first.

The minimum energy required is reported in Table 19.9.

19.5.2.4 Tire Endurance

The test is basically a fatigue-damaging process (it differs from the high-speed test of the ECE standard (Section 19.5.1), which is an integrity test at the maximum speed).

The tire is mounted on the rim and inflated according to standards [27] and maintained into a room at 100°F ± 5°F (38°C ± 3°C). The tire is pressed against a wheel with a diameter of 67.23 in. (1708 mm). The test velocity is imposed at 120 km/h, and the load is set accordingly to what specified in Table 19.10.

The test is passed if the pressure is not dropped and if the tire meets all visual inspection criteria, that is, no visual evidence of tread, sidewall, ply, cord, inner liner, belt, or bead separation, chunking, open splices, cracking, or broken cords.

TABLE 19.10
Loads to Be Applied for the DOT Tire Endurance Test

Testing Hours (h)	Percent of Load Rating (%)
4	85
6	90
24	100

19.5.2.5 Low-Pressure Test

This test is conducted following completion of the tire endurance test (see previous section), using the same tire and rim assembly. The tire has to be deflated to the appropriate pressure reported in [27] (for passenger car tires, the pressure can be either 1.4 or 1.6 bar).

The test conditions (such as temperature) are similar to the endurance test ones. The test lasts 90 min at a speed of 120 km/h at 100% of load capacity. Between 15 and 25 min after the completion of the 90 min test, the pressure should be not less than 95% of the initial pretest pressure. Other visual tests are to be performed.

19.5.2.6 High-Speed Performance

The test conditions (such as temperature) are similar to the endurance test ones. The tire is mounted on the rim and inflated according to standards [27]. The tire is pressed against a wheel with diameter of 67.23 in. (1708 mm). The load is set to 85% of the maximum load capacity.

A break-in cycle shall be conducted for at least 2 h at 80 km/h (50 mph). At the conclusion of the run, unloading the tire will be allowed to cool.

The main test will commence within 30 min after the cooling cycle. All test cycles (140, 150, and 160 km/h or 87, 93, and 99 mph) will be conducted for 30 min each, without interruption.

The tire passes the tests if the pressure does not drop and no damages can be detected as in the previous tests.

19.5.3 Internal Experimental Tests

Tests have to be performed in the whole process of tire production, starting from the design phase and during the current production (quality control).

Inside each company, specific (and severe) standards are defined and used, and many of them are statements on the way the tests have to be performed. Such tests can be classed as described in Figure 19.30.

19.5.3.1 Type A Tests

The typical standards provide information on how to evaluate the structural integrity of tires (Figure 19.31).

The fatigue tests are carried out on a drum, under specific conditions of pressure, speed, vertical load, and external temperature. The aim is to find the number of revolutions before a failure occurs. Failures at the end of the test can be either the separation of the belt edges, due to shear action, or disjunctions at the bead area. The test parameters are tuned to reproduce the desired kind of failure in the target market.

19.5.3.2 Type B Tests

The handling test is not dealt with in this chapter; see other chapters in this book. The comfort issues are dealt with in the sequel.

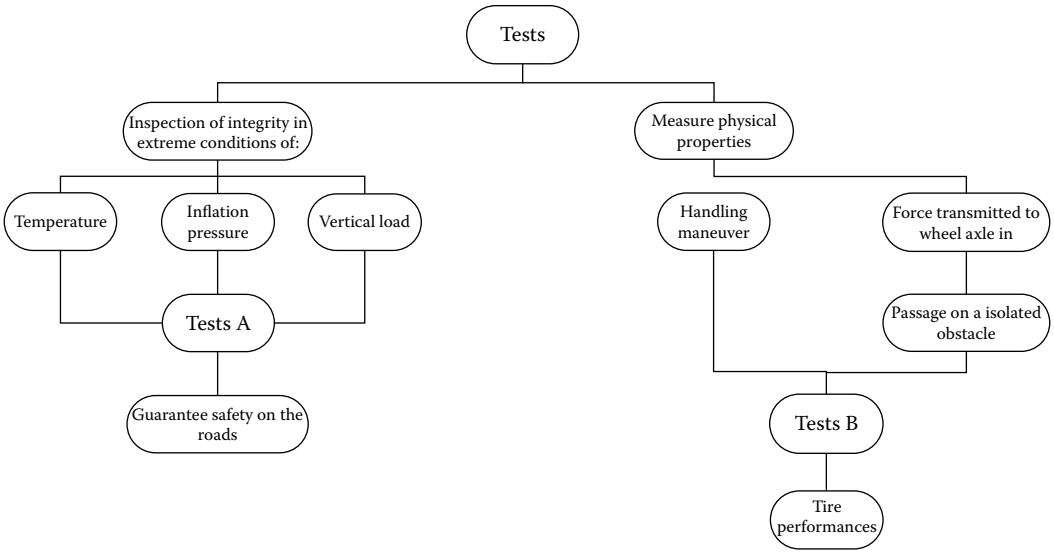


FIGURE 19.30 Typical internal experimental tests.

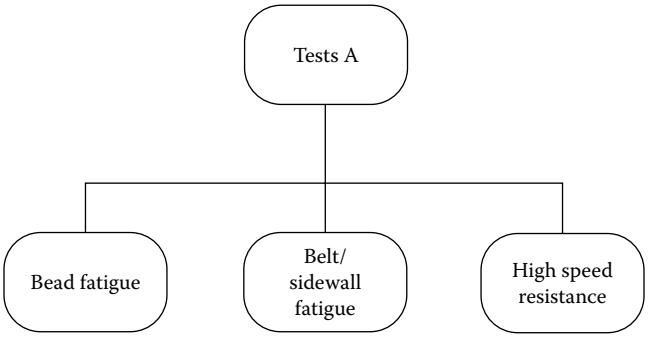


FIGURE 19.31 Type A tests.

19.5.3.2.1 Comfort Tests

A technique for determining the parameters of a rigid ring model of the tire is presented. The model can be used for studying comfort and braking issues in the 0–100 Hz.

The technique used to determine the parameters of the model includes both the direct measurements (mass, moment of inertia, stiffness) and a method to identify some parameters.

The tire is modeled as a rigid ring [28] (Figure 19.32).

The model is described by the following parameters:

a. Mass m_b and moment of inertia J_b of the rigid ring (mass and moment of the tread, belt pack, and carcass)
b. A moment of inertia J_r of the wheel, which includes a part of the tire moment of inertia (bead)
c. "Effective rolling radius" r_e of the tire, equal to radius R of the rigid ring
d. Radial c_b and torsional c_{bt} "structural stiffness": springs that represent, respectively, the pneumatic stiffness of the sidewall and the carcass, with the tire inflated
e. Vertical c_{cz}, introduced to take into account the deformations in the contact area (patch)

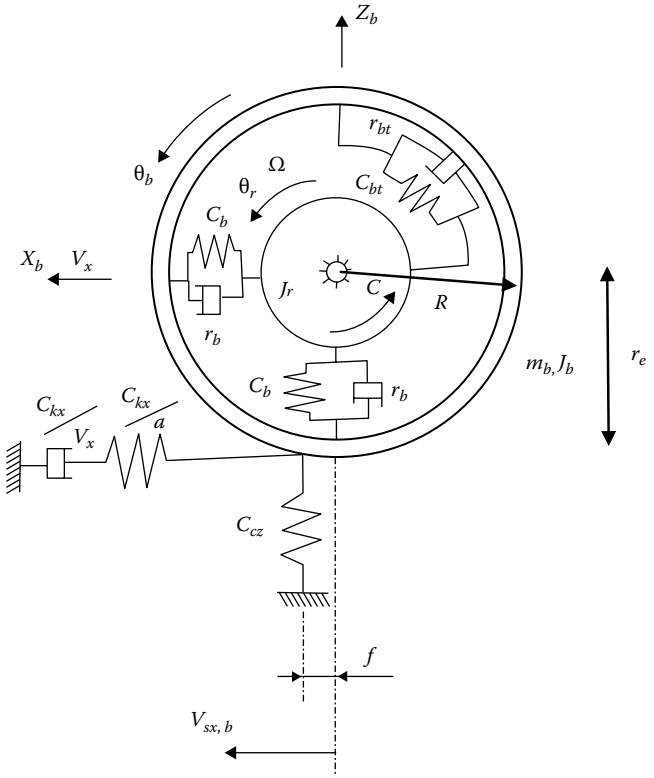


FIGURE 19.32 Rigid ring model.

f. Damping parameters corresponding to the stiffnesses indicated at points (d) and (e): r_b, r_{bt}
g. A "first-order model" (damper plus spring in series) taking into account the longitudinal slip in the contact patch. The following parameters are defined:
   - Longitudinal slip stiffness c_{kx} (= $2a^2C_p$, where C_p is the stiffness of the tread per unit length)
   - Length of the contact patch $2a$
   - Forward wheel velocity V_x
   - Slip velocity V_{sx}, b

The equations of motion concerning the model behavior are reported in [30].

The parameters to be measured directly are the following:

$$m_b;\ J_b;\ J_r;\ c_{bt};\ r_{bt};\ c_{kx};\ r_e;\ a$$

the parameters to be identified are

$$C_b,\ r_b,\ C_{cz}$$

19.5.3.2.1.1 Measuring the Length of the Contact Patch 2a Given that the contact patch of a tire varies with the forward velocity (the measurement performed with a stationary tire is much different from the measurement executed while it is rolling), the measurement is performed by photographing a tire mounted on a vehicle as it passes over a "glass plate." In particular, the camera is located underneath the road that is covered with a glass plate. Typical values of the length $2a$ (under a static load) are in the range 50–150 mm.

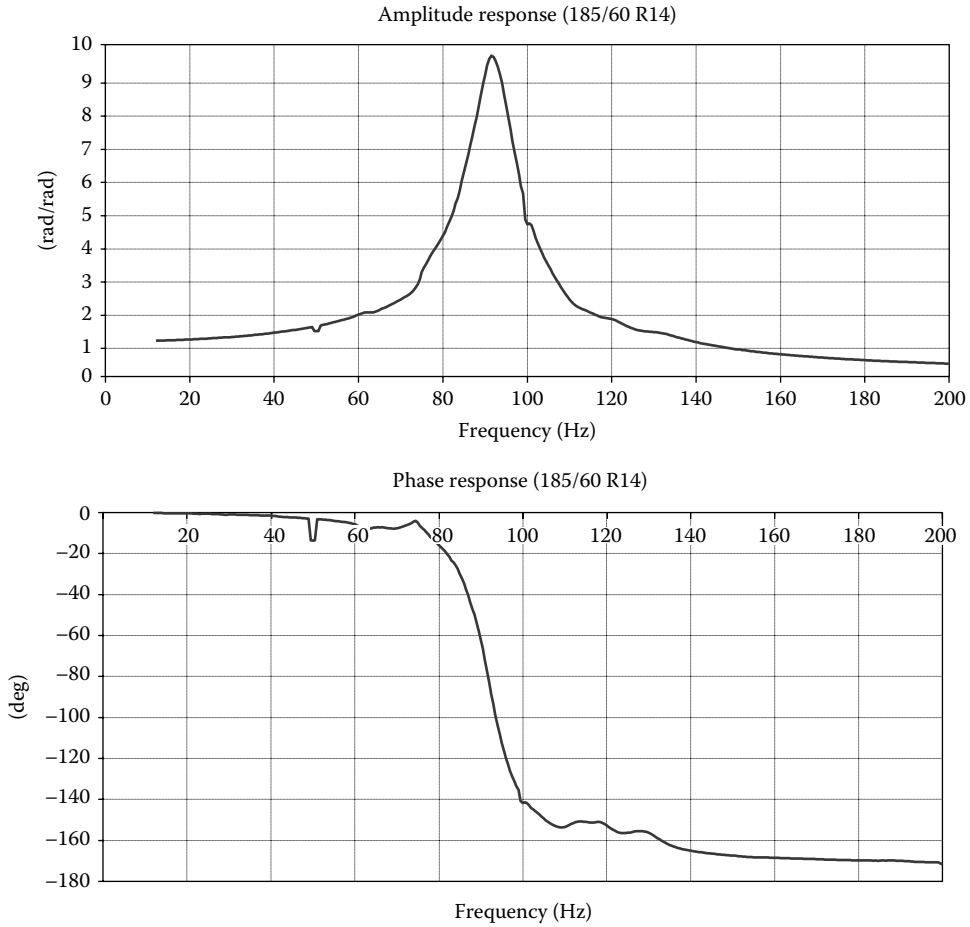


FIGURE 19.33 Experimental transfer function $H^{(s)}(\Omega)$ of a car tire.

19.5.3.2.1.2 Measuring the Mass Parameters A torsional pendulum [30] is used to measure the moment of inertia of the wheel J_{TOT} (rim plus tire). The moment of inertia of the tire J_p is determined by

$$J_p = J_{TOT} - (J_m + J_c)$$

where

J_m is the moment of inertia of the hub
J_c is the moment of inertia of the rim

Typical values of J_p and J_c (car tires) are, respectively, in the ranges 0.4–1.2 kg m^2 and 0.15–0.35 kg m^2.

The moment of inertia J_p of the tire can be divided, basically, into a moment of inertia of the tread rubber + belts + casing J_b and a moment of inertia related to the beads. Usually, $J_b = 2/3\ J_p$.

In the model in Figure 19.33, J_r has to account for the moment of inertia of the bead.

The parameter m_b (mass of the rigid ring) is $m_b = J_b/R^2$ ($r_e = R$).

19.5.3.2.1.3 Measuring the Stiffness c_{bt} and Damping r_{bt} An electric exciter is used to measure c_{bt} and r_{bt}. The exciter generates an alternated rotation of the rim on which the tire is mounted. The respective angular acceleration of the rim a_C and of the tire a_B (i.e., the tread + belts + carcass)

is measured. The ratio between the two accelerations as function of the excitation frequency Ω provides the *experimental* transfer function $H^{(s)}(\Omega)$

$$H^{(s)}(\Omega) = \frac{a_B(\Omega)}{a_C(\Omega)}$$

A typical transfer function $H^{(s)}(\Omega)$ is reported in Figure 19.33.

It should be noted that $H^{(s)}(\Omega)$ can also be expressed as

$$H^{(s)}(\Omega) = \frac{-\Omega^2 \vartheta_B}{-\Omega^2 \vartheta_C} = \frac{\vartheta_B}{\vartheta_C}$$

where ϑ_B and ϑ_C are the *measured* rotations, respectively, of the tread and the rim.

Considering the moment of inertia of the rigid ring J_b connected to the moment of inertia of the rim J_r through the torsional spring and dashpot c_{bt} and r_{bt}, the *calculated* transfer function $h^{(c)}(\Omega)$ between the calculated rotation of the tread ϑ_b and calculated rotation of the rim ϑ_r is given by

$$h^{(c)}(\Omega) = \frac{\vartheta_b}{\vartheta_r} = \frac{i\Omega r_{bt} + c_{bt}}{-\Omega^2 J_b + i\Omega r_{bt} + c_{bt}}$$

By assigning the proper value to J_b, it is possible to identify the parameters r_{bt} and c_{bt}.

Typical values of (car tire) r_{bt} and c_{bt} are found, respectively, in the ranges 18–25 N-m s/rad and 80×10^3–150×10^3 N-m/rad.

19.5.3.2.1.4 Measuring the Sliding Stiffness c_{kx} The parameter c_{kx} represents the longitudinal stiffness of the tread in the contact patch. This parameter depends on the compound and the tread pattern, as well as the extension of the tire–road contact area (and therefore the vertical load). A dynamometric trailer can be used to measure the longitudinal stiffness. Alternatively, the tire can be tested on a flat belt machine. A longitudinal slip is applied $V_{sx} = V_x - r_e\Omega$ (see Figure 19.34) and the longitudinal force F_x is measured. For small V_{sx}, $c_{kx} = \Delta F_x/\Delta V_{sx}$. This procedure is similar to the one reported in [29].

Typical values of c_{kx} are found in the range 20,000–150,000 N, and they depend strongly on the vertical load.

19.5.3.2.2 Tire Cleat Test

The natural frequencies and the damping values of the tire in the vertical plane are derived, in order to identify C_b, r_b, C_{cz}. The test involves a wheel passing over an obstacle. The wheel is mounted on a fixed measuring hub and rolls on a drum [30]. The signals—as function of time—measured at the hub refer to both the longitudinal and vertical forces, which are generated by the excitation imposed by the obstacle. By applying a proper signal analysis [30], the values of the frequencies and the damping of the main plane vibrating modes in the range 30–130 Hz are estimated. Such modes are usually four (see [29,30]) and depend on the forward velocity (Figure 19.35). Finally, the parameters C_b, r_b, and C_{cz} are computed as shown in [30] by exploiting the equations of motion of the system model in Figure 19.32.

Typical values of parameters C_b, r_b, and C_{cz} are found, respectively, in the ranges 8×10^5–20×10^5 N/m, 200–500 N s/m, and 100×10^3–300×10^3N/m.

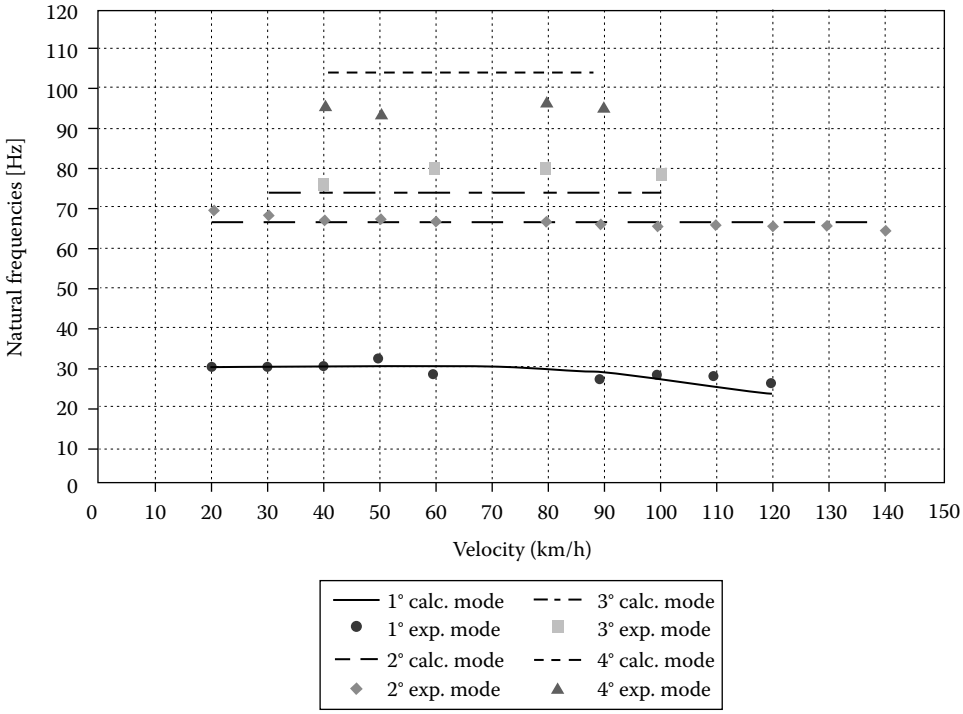


FIGURE 19.34 Measured and computed in plane natural frequencies of a rolling tire. The computations are performed by the model in Figure 19.32.

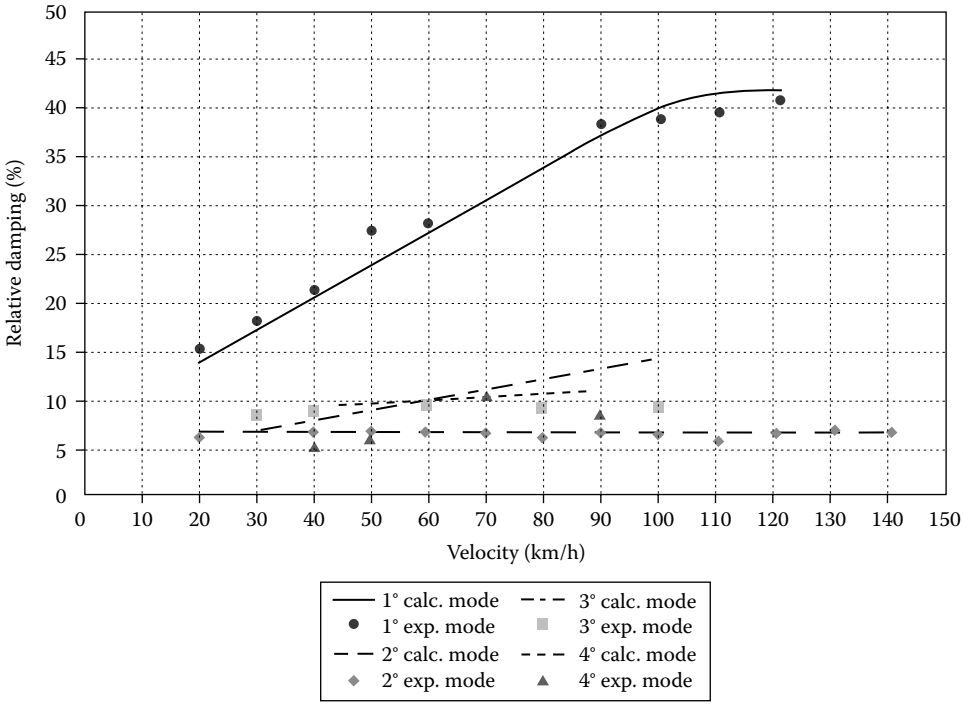


FIGURE 19.35 Measured and computed damping (r_b) of a rolling tire. The computations are performed by the model in Figure 19.32.

19.5.4 Other Standards or Regulations

Additionally, with respect to the historical regulations and standards reported in the previous sections, the following regulations or standards have been issued or are being issued at the time this chapter is written:

- ISO 4000: The standard is divided into two parts. The first part specifies the designation, dimensions, and load ratings of metric–series tires primarily intended for passenger cars. The second part refers to rims.
- ISO 28580: The standard specifies methods for measuring rolling resistance, under controlled laboratory conditions, for new pneumatic tires designed primarily for use on passenger cars, trucks, and buses. Tires intended for temporary use only are not included. The standard includes a method for correlating measurement results to allow interlaboratory comparisons. Measurement of tires using this method enables comparisons to be made between the rolling resistance of new test tires when they are free-rolling straight ahead, in a position perpendicular to the drum outer surface, and in steady-state conditions. In measuring tire rolling resistance, it is necessary to measure small forces in the presence of much larger (vertical) forces. It is, therefore, essential that equipment and instrumentation of appropriate accuracy be used.
- ISO 15222: The standard specifies the method for measuring the relative wet grip braking performance index to a reference under loaded conditions for new tires for use on commercial vehicles on a wet-paved surface.
- REGULATION EC 661/2009: The regulation concerns type-approval requirements for the general safety of motor vehicles, their trailers and systems, components, and separate technical units intended therefore. Requirements for tires with regard to wet grip, rolling resistance, and rolling noise are included. The regulation refers to E/ECE/324/Rev.2/Add.116/Rev.2, E/ECE/TRANS/505/Rev.2/Add.116/Rev.2 Regulation No. 117, where uniform provisions concerning the approval of tires with regard to rolling, to sound emissions, and to adhesion on wet surfaces and/or to rolling resistance are reported.
- ASTM 1805: The standard refers to a test method for single wheel driving traction in a straight line on snow- and ice-covered surfaces.

19.6 CONCLUSIONS

The tire is a fascinating complex device. After more than 100 years the tire has been invented, there are still huge theoretical and experimental gaps to be covered. In the future, a continuous evolution toward cheaper and more effective processes is expected. New materials could dramatically change both the tire production and the tire performance.

REFERENCES

1. Mortensen, A., *Concise Encyclopedia of Composite Materials*, Elsevier, Amsterdam, the Netherlands, 2007.
2. www.pirelli.com
3. Clarks, S. (ed.), *Mechanics of Pneumatic Tyres*, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, 1980.
4. www.aipma.net/info/plasticprocess.htm
5. *The Tyre Digest*, Edouard Michelin, Watford, U.K., 2002.
6. www.iisrp.com
7. Walter, J.D., Gent, A. (eds.), *The Pneumatic Tyre*, NHTSA, Washington, DC, 2005.
8. www.carbon-black.org
9. Donnet, J.B. et al., *Carbon Black Science and Technology*, Marcel Dekker, New York, 1993, ISBN 082478975X, 9780824789756.

10. www.etrma.org
11. Donnet, J.B., Voet, A., *Carbon Black: Physics, Chemistry, and Elastomer Reinforcement*, Marcel Dekker, New York, 1976, ISBN 0824763459, 9780824763459.
12. www.p2pays.org/ref/11/10504/html/intro/tire.htm
13. www.emt-india.net/process/tyre/tyre.htm
14. Alliger, G., Sjothun, I.J., *Vulcanization of Elastomers*, Reinhold Publishing Co., New York, 1963.
15. Niyogi, U., *Polymer Additives and Compounding*. Shri Ram Institute for Industrial Research, Delhi, India, 2007.
16. Scheiderich, R.F. et al., Method for preparing tire curing bladder lubricant. U.S. patent 4509984.
17. James, M., Burak, E., *Science and Technology of Rubber*, Elsevier, San Diego, CA, 2005.
18. www.us-tra.org
19. www.etrto.org
20. Regulation (EC) No 1222/2009 of the European Parliament and of the Council of 25 November 2009 on the labelling of tyres with respect to fuel efficiency and other essential parameters.
21. Commission Regulation (EU) No228/2011 of 7 March 2011 amending Regulation (EC) No1222/2009 of the European Parliament and of the Council with regard to the wet grip testing method for C1 tyres.
22. www.unece.org
23. United Nations, Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers, UN/ECE Regulation 30, E/ECE324, 2007.
24. ISO 4000 Passenger car tyres and rims.
25. www.dot.gov
26. www.nhtsa.gov
27. NHTSA, Laboratory test procedure for FMVSS no. 139, New pneumatic radial tires for light vehicles, Washington, DC, January 2006.
28. Van Oosten, J., Pacejka, H.B., SWIFT-tyre: An accurate tyre model for ride and handling studies also at higher frequencies and short road wavelengths, *ADAMS User Conference*, Orlando, FL, June 2000.
29. Zegelaar, P.W.A., Pacejka, H.B., The in-plane dynamics of tyres on uneven roads, *Vehicle System Dynamics*, *14th IAVSD Symposium of Vehicles on Roads and Tracks*, Ann Arbor, MI, August 21–25, 1995.
30. Mancosu, F., Matrascia, G., Cheli, F., Techniques for determining the parameters of a two-dimensional tire model for the study of ride comfort, *Tire Science and Technology, TSTCA*, 25(3), 187–213, July–September, 1997.
31. www.avontyres.com.au/technical/images/cutaway-passenger.jpg
32. www.2405.com/press-library/Hummer-H3-2006/cutaway-of-the-bridgestone-lt285-75r16c-33-inch-on-off-road-radial-tire.jpg
33. www.corryrubber.com/main/ns/67/category/7/doc/56
34. French, T., *Tyre Technology*, Adam Hilger, Bristol, U.K., 1989.
35. Mark, J. et al., *Science and Technology of Rubber*, Academic Press, New York, 1994.
36. www.sae.org
37. SAE, Passenger car tire performance requirements and test procedures, J918, 1970.
38. www.toyo.com/au
39. Jodry, C., Boynton, R., Determining product of inertia using a torsion pendulum, *41st Annual Conference of the Society of Allied Weight Engineers, Inc*., San José, CA, May 17–19, 1982.
40. ASTM, Standard classification for carbon blacks used in rubber production, ASTM D1765, 2000.
41. ISO, Rubber—Determination of crystallization effects by hardness measurements, ISO 3387, 1994.
42. www.tirelitigation.com

20 Mechanics of Off-Road Vehicle–Terrain Interaction
Terramechanics

Jo Y. Wong

CONTENTS

20.1 Introduction 697
20.2 Applications of the Theory of Elasticity to the Analysis of Stress Distribution in the Terrain under Vehicular Loads 698
20.3 Applications of the Theory of Plastic Equilibrium to the Prediction of Vehicle Traction 702
20.4 Empirical Methods for Predicting Off-Road Vehicle Performance Based on the Cone Index 705
20.4.1 Empirical Method for Predicting Tracked Vehicle Performance 706
20.4.2 Empirical Method for Predicting Wheeled Vehicle Performance 708
20.5 Empirical Methods for Predicting Vehicle Mobility Based on the Mean Maximum Pressure 709
20.6 Semiempirical Methods for Predicting the Performance of Tracked Vehicles 711
20.6.1 Bevameter Techniques 711
20.6.1.1 Pressure–Sinkage Relationships 712
20.6.1.2 Shear Stress–Shear Displacement Relationships 713
20.6.1.3 Response of Terrain to Repetitive Loading 714
20.6.2 Motion Resistance of a Track 715
20.6.3 Thrust–Slip Relationship of a Track 716
20.7 Semiempirical Methods for Predicting the Performance of Wheeled Vehicles 718
20.7.1 Motion Resistance of a Rigid Wheel 718
20.7.2 Motion Resistance of a Pneumatic Tire 719
20.7.3 Thrust–Slip Relationship of a Wheel 720
20.8 Closing Remarks 722
Acknowledgments 723
References 723

20.1 INTRODUCTION

While transportation over paved roadways and railways is of vital importance to the economy of a developed society, locomotion over unprepared terrain still plays a significant role in a wide range of human endeavors, including farming, construction, logging, mining, exploration, recreation, cross-country transportation, and military operations. For a long period of time in the past, owing to the lack of understanding of the behavior of the terrain and of the mechanics of vehicle–terrain interaction, the development and design of off-road vehicles have been, by and large, guided by past experience and the "cut and try" methodology. In recent years, with the increasing globalization of the economy, improving the competitiveness in the world market has become a major focus of industry.

The empirical approach to the development of off-road vehicles adopted in the past has become extremely inefficient and prohibitively expensive. This has led to the recognition of the necessity of systematic studies of the principles underlying the rational development and design of off-road vehicles. The study of the performance of off-road vehicles and machinery in relation to their operating environment, the terrain, has now become known as *terramechanics* [1–4].

In off-road operations, various types of terrain, ranging from desert sand through soft mud to fresh snow, may be encountered. The properties of the terrain quite often impose severe limitations to the mobility of off-road vehicles. An adequate knowledge of the mechanical properties of the terrain and an understanding of its response to vehicular loading are, therefore, essential to the rational development and design of off-road vehicles to meet specific operational requirements. The role of terramechanics in the development of off-road vehicles is, perhaps, analogous to that of aerodynamics in the development of aircraft or that of hydrodynamics in the design of marinecraft.

On a given terrain, the performance of an off-road vehicle is to a great extent dependent upon the manner in which the vehicle interacts with the terrain. Consequently, an understanding of the mechanics of vehicle–terrain interaction is of importance to the rational development and design of off-road vehicles. The central issue is to establish a quantitative relationship between the performance and design of an off-road vehicle for a given operating environment. In this chapter, various methods for the study of vehicle–terrain interaction are briefly reviewed.

20.2 APPLICATIONS OF THE THEORY OF ELASTICITY TO THE ANALYSIS OF STRESS DISTRIBUTION IN THE TERRAIN UNDER VEHICULAR LOADS

Certain types of terrain, such as dense soil, may be idealized as an elasto-plastic material. When the stress level in the terrain does not exceed a certain limit, it may be assumed to exhibit elastic behavior. The idealization of the terrain as an elastic medium has found applications in the prediction of stress distribution in the soil, in connection with the study of soil compaction due to vehicular loading [5].

The prediction of stress distribution in an elastic medium subject to any type of loading on the surface may be based on the analysis of the distribution of stresses under a point load. A method was developed by Boussinesq for predicting the stress distribution in a semi-infinite, homogeneous, isotropic, elastic medium subject to a vertical point load on the surface [1,3]. It gives the following expressions for the vertical stress σ_z at a point defined by the coordinates shown in Figure 20.1:

$$\sigma_z = \frac{3W}{2\pi} z^3 (x^2 + y^2 + z^2)^{-5/2} = \frac{3W}{2\pi[1 + (r/z)^2]^{5/2} z^2}$$

$$= \frac{3W}{2\pi R^2} \left(\frac{z}{R} \right)^3 = \frac{3W}{2\pi R^2} \cos^3 \theta \tag{20.1}$$

where

$R = (x^2 + y^2 + z^2)^{1/2}$

$\cos \theta = z/R$

When polar coordinates are used, the radial stress σ_r is given by

$$\sigma_r = \frac{3W}{2\pi R^2} \cos \theta \tag{20.2}$$

It is interesting to note that the stresses are independent of the modulus of elasticity of the material and that they are only functions of the load applied and the distance from the point of application of the load. It should be mentioned that Equations 20.1 and 20.2 are only valid for predicting stresses at points not too close to the point of application of the load. The material in the immediate vicinity of the point load does not behave elastically.

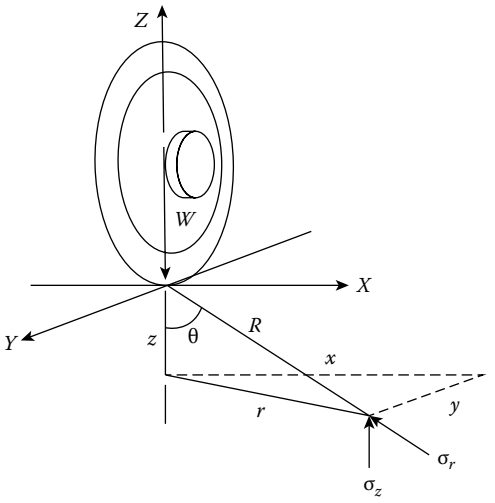


FIGURE 20.1 Stresses at a point in a semi-infinite elastic medium subject to a point load on the surface. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956, with permission of the University of Michigan Press.)

It should be pointed out that the use of the theory of elasticity for predicting stresses in the field produces approximate results only. Measurements have shown that the stress distribution in the soil deviates from that computed using Boussinesq's equation [5]. There is a tendency for the compressive stresses in the soil to concentrate around the loading axis. This tendency becomes greater when the soil becomes more plastic due to increased moisture content. In view of this, various semiempirical equations have been proposed to account for the different characteristics of various types of soil. Frohlich introduced a concentration factor v to Boussinesq's equations. The factor v reflects the behavior of different types of soil under different conditions. Introducing the concentration factor v, the expressions for the vertical and radial stresses in the soil subject to a point load on the surface take the following forms:

$$\sigma_z = \frac{vW}{2\pi R^2}\cos^v \theta \tag{20.3}$$

$$\sigma_r = \frac{vW}{2\pi R^2}\cos^{v-2} \theta \tag{20.4}$$

The value of the concentration factor v depends on the type of soil and its moisture content. It varies from 4 for hard, dry soil to 6 for wet soil. For fairly moist and relatively dense soil, the value of v is 5. It should be noted that for an elastic medium, the value of v is 3 and that Equations 20.3 and 20.4 become identical to Equations 20.1 and 20.2, respectively.

Based on the analysis of stress distribution under a point load, the distribution of stresses in a soil under a variety of loading conditions may be predicted using the principle of superposition [1,3]. For instance, for a circular contact area with a radius r_o and a uniform contact pressure p_o shown in Figure 20.2, the vertical stress at a depth z below the center of the contact area may be determined by considering the load applied on the contact surface consisting of a number of discrete point loads. Each of these point loads is $dW = p_o\, dA = p_o\, rdrd\theta$. Hence, in accordance with Equation 20.1, for a terrain exhibiting elastic behavior with $v = 3$,

$$d\sigma_z = \frac{3p_o rdrd\theta}{2\pi[1+(r/z)^2]^{5/2}z^2} \tag{20.5}$$

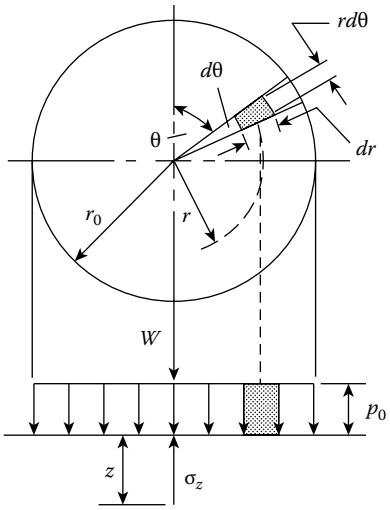


FIGURE 20.2 Vertical stresses in a semi-infinite elastic medium below the center of a circular loading area with uniform contact pressure. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956, with permission of the University of Michigan Press.)

The resultant vertical stress σ_z at a depth z below the center of the circular contact area is then equal to the sum of the stresses produced by the point loads, each of which is $p_0\, rdrd\theta$. It can be computed by a double integration:

$$\sigma_z = \frac{3p_0}{2\pi}\int_0^{r_0}\int_0^{2\pi}\frac{rdrd\theta}{[1+(r/z)^2]^{5/2}z^2} = p_0\left[1-\frac{z^3}{(z^2+r_0^2)^{3/2}}\right] \tag{20.6}$$

For a terrain with value of v other than 3, the resultant vertical stress below the center of the contact area can be predicted following a similar approach to that described earlier. For instance, when $v = 4$, the resultant vertical stress σ_z at a depth z below the center of the circular contact area is expressed by [3]

$$\sigma_z = p_0\left[1-\frac{1}{\left[1+\left(r_0/z\right)^2\right]^2}\right] \tag{20.7}$$

The previously noted approach can be applied to the prediction of the stress distribution in the soil under a tire, if the contact area of the tire and the pressure distribution over the contact surface are known. Figure 20.3 shows the distributions of the major principal stress in a soil having normal field density and moisture content and with $v = 5$ under tires of different sizes and normal loads but with the same tire inflation pressure [5]. It shows that at the same tire inflation pressure, the stress can penetrate deeper with larger tires carrying heavier loads. This is because the larger tire has a larger contact area, and as a result, the stress at the same depth increases as shown in Figure 20.3, although the pressure applied on the surface of the soil remains the same. Figure 20.4 shows the effects of soil conditions on the stress distribution in the soil under a tire. In hard, dry, and dense soil with $v = 4$, the lines of equal major principal stress, commonly known as stress bulbs, are approximately circular. For wet soil with $v = 6$, the stress is more concentrated toward the center of the loading area. For tires with lugs, such as tractor tires, the pressure distribution over the contact area is dependent on soil

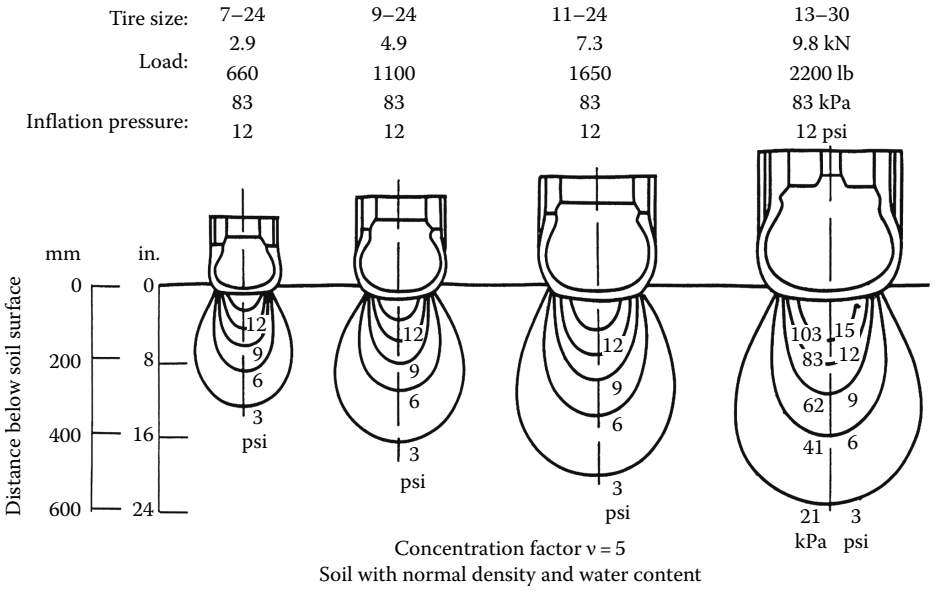


FIGURE 20.3 Distribution of major principal stresses in a soil under the loading of various tires. (Reproduced from Sohne, W., Fundamentals of pressure distribution and soil compaction under tractor tires, *Agr. Eng.*, 39(5), 276–281, 1958. With permission of the American Society of Agricultural Engineers.)

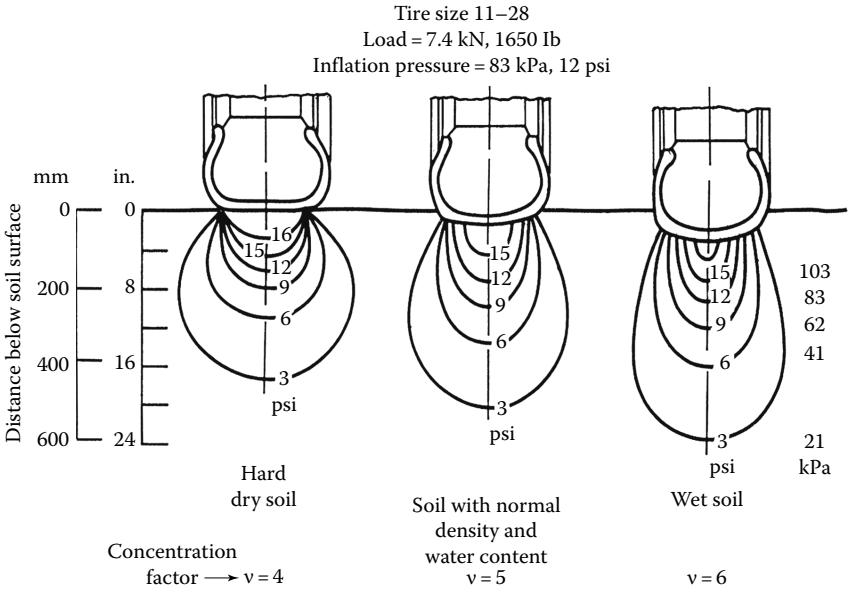


FIGURE 20.4 Distribution of major principal stresses under a tire at different soil conditions. (Reproduced from Sohne, W., Fundamentals of pressure distribution and soil compaction under tractor tires, *Agr. Eng.*, 39(5), 276–281, 1958. With permission of the American Society of Agricultural Engineers.)

conditions. In hard soil, the lugs of the tire may carry the entire load. The pressure over the contact area of the lugs may be three to four times higher than that of an equivalent tire without lugs. In wet soil, because of the sinkage of the tire, there may be hardly any difference between the contact pressure under the lugs and that under the carcass. In principle, the stress distribution in the soil under tires with lugs may be estimated following an approach similar to that described earlier.

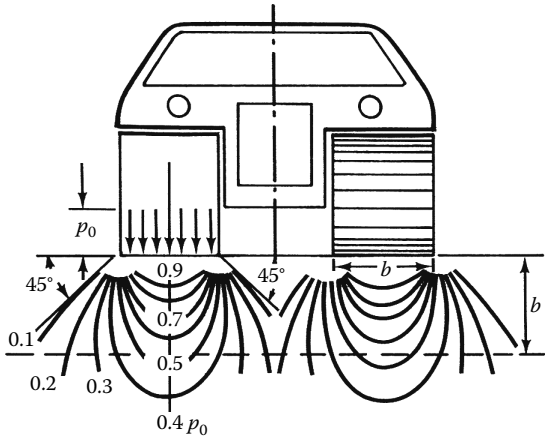


FIGURE 20.5 Distribution of vertical stresses in a semi-infinite elastic medium under a tracked vehicle. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956, with permission of the University of Michigan Press.)

The stress distribution in the soil under a tracked vehicle may be predicted in a similar way. Figure 20.5 shows the distribution of vertical stress in a soil with $v = 3$ (i.e., the soil is assumed to be elastic) under the loading of a tracked vehicle [1,3]. In the analysis, the track is idealized as a rigid footing with a high ratio of track length to width and with a uniform contact pressure. It is interesting to note from Figure 20.5 that at a depth equal to the width of the track, the vertical stress under the center of the track is approximately 50% of the pressure applied on the contact surface and that it practically vanishes at a depth equal to twice the width of the track. The boundaries of the stress bulbs, for all practical purposes, may be assumed as being sloped at an angle of 45° with the horizontal, as shown in Figure 20.5. Outside the boundaries, the stress in the soil is negligible.

It should be pointed out that in practice the applications of the theory of elasticity to the analysis of vehicle–terrain interaction are limited, as it is applicable only to dense soil where soil deformation is small.

20.3 APPLICATIONS OF THE THEORY OF PLASTIC EQUILIBRIUM TO THE PREDICTION OF VEHICLE TRACTION

When the vehicular load applied to the terrain surface exceeds a certain limit, the stress level within a certain boundary of the terrain may reach a threshold that will initiate plastic flow of the terrain material. The state that precedes plastic flow is usually referred to as plastic equilibrium. The transition from the state of plastic equilibrium to that of plastic flow represents the failure of the terrain material.

There are a number of proposed criteria for the failure of soils and other similar materials. A widely used criterion is that due to Mohr–Coulomb. It postulates that the material at a point will fail if the shear stress at that point satisfies the following condition [1,3,4]:

$$\tau = c + \sigma \tan \phi \tag{20.8}$$

where

τ is the shear strength of the material
c is the apparent cohesion
σ is the normal stress
ϕ is the angle of internal shearing resistance of the material

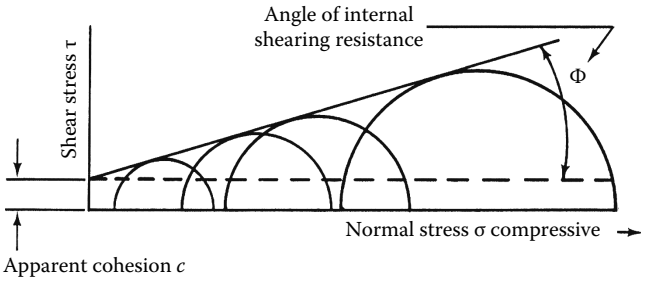


FIGURE 20.6 Mohr–Coulomb failure criterion. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

Cohesion of the material is the bond that cements particles together irrespective of the normal pressure exerted by one particle upon the other. On the other hand, particles of frictional material can be held together only when a normal pressure exists between them. Thus, the shear strength of saturated clay and the like does not depend on the normal load, whereas the shear strength of dry sand and the like increases with an increase in the normal pressure. For most soils, they usually exhibit both cohesive and frictional behavior.

The meaning of the Mohr–Coulomb failure criterion may be illustrated with the aid of the Mohr circle of stress, as shown in Figure 20.6. If specimens of a soil are subject to different states of stress, for each mode of failure, a Mohr circle can be constructed. If a straight line enveloping the set of Mohr circles is drawn, it will be in the form of Equation 20.8, with the cohesion of the soil being determined by the intercept of the enveloping straight line with the shear stress axis and the angle of internal shearing resistance being represented by its slope. The Mohr–Coulomb failure criterion is simply that if a Mohr circle representing the state of stress at a point in the soil touches the envelope shown in Figure 20.6, failure will take place at that point.

The shear strength parameters c and ϕ may be measured by a variety of devices [3,4]. The triaxial apparatus and the translational shear box are commonly used in civil engineering. For vehicle mobility study, however, annular shear plates (shear rings) are usually used to obtain the shear strength of terrain in the field [3,4].

The Mohr–Coulomb failure criterion can be directly applied to the prediction of the maximum thrust that an off-road vehicle can develop on a given terrain. It should be noted that the development of thrust is a result of the vehicle running gear shearing the terrain surface, as shown in Figure 20.7. For instance, if a tracked vehicle has a total track contact area A and a uniform contact pressure σ, then using Equation 20.8, the maximum thrust F_{max} of the tracked vehicle may be estimated as follows:

$$F_{max} = \tau A = (c + \sigma \tan \phi)A = cA + W \tan \phi \tag{20.9}$$

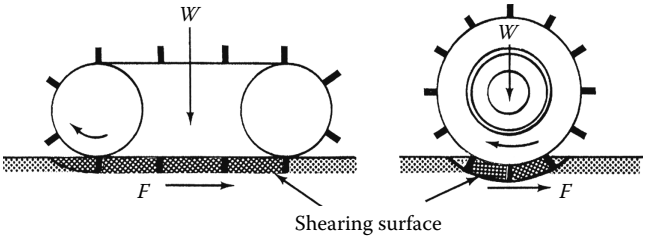


FIGURE 20.7 Shearing action of a track and a wheel. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

where W is the vehicle weight that is equal to the uniform normal pressure σ multiplied by the total track contact area A.

From Equation 20.9, it is seen that on a cohesive terrain with negligible angle of internal shearing resistance, such as wet clay, the maximum thrust of the vehicle is primarily dependent on the total track contact area and the cohesion of the terrain, while vehicle weight has little effect on the development of thrust. On the other hand, on dry desert sand where cohesion is negligible, the maximum thrust will be primarily dependent upon the vehicle weight and the angle of internal shearing resistance of the sand, while track contact area has little effect on the thrust.

Another important application of the theory of plastic equilibrium is the development of the theory of passive earth pressure [1,3]. It can be used in the prediction of the force acting on a bulldozer blade or of the thrust developed by a lug (grouser) of a wheel, as shown in Figure 20.8 [3]. Based on the theory of passive earth pressure, the force F_p acting on a vertical bulldozer blade with a smooth surface may be calculated as follows [3]:

$$F_p = b\int_0^{h_b} \sigma_p dz = b\int_0^{h_b} \left(\gamma_s z N_\phi + 2c\sqrt{N_\phi}\right) dz$$

$$= b\left(\frac{1}{2}\gamma_s h_b^2 N_\phi + 2ch_b\sqrt{N_\phi}\right) \tag{20.10}$$

where

b is the width of the bulldozer blade
σ_p is the passive earth pressure
γ_s is the weight density of the terrain
h_b is the cutting depth of the blade
N_ϕ is called flow value that is equal to $\tan^2(45° + \phi/2)$
c is the cohesion
ϕ is the angle of internal shearing resistance of the terrain

A similar approach may be followed to estimate the thrust developed by a lug of a cage wheel, such as that used in paddy field or that attached to a tire as a traction-aid device used in wet soil. It should be noted, however, that the lug may behave in one of two ways. If the spacing between the lugs is too small that the space between them is filled up with soil, then the shearing of the soil takes place across the lug tips. Under these circumstances, the major effect of the lugs would be the increase of the effective diameter of the wheel. On the other hand, if the spacing between the lugs is large enough that the soil fails in the manner shown in Figure 20.8, then the behavior of the lug

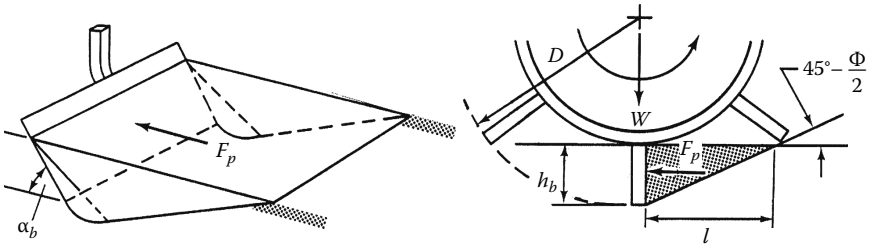


FIGURE 20.8 Interaction of a soil cutting blade and a grouser of a wheel with soil. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

will be similar to that of a bulldozer blade. In this case, the thrust developed by a lug with a smooth surface in the vertical position may be predicted using Equation 20.10. It should be mentioned that the shearing forces developed on the vertical surfaces on both sides of the lug may increase the total thrust and that they should be taken into account when the height of the lug is large and its penetration into the soil is significant [3,4].

It should be pointed out that in practice, a bulldozer blade is usually not vertical and the angular position of a lug on a wheel varies as it rotates. Furthermore, the bulldozer blade or the lug surface is usually not smooth and there is shear stress acting on the interface. As a result, the interaction between an inclined bulldozer blade or a lug at various angular positions and the terrain is more complicated than that described earlier, and Equation 20.10 does not provide accurate predictions of the interacting forces. Methods for predicting the forces acting on an inclined bulldozer blade with non-smooth contact surface or a lug with rough surface at various angular positions are described in Ref. [3].

It should also be noted that strictly speaking the theory of plastic equilibrium is only applicable to plastic material with little compressibility. In reality, natural terrain has varying degrees of compressibility and does not behave like a perfectly plastic material. Also, the theory of plastic equilibrium can only be used to predict the loads or forces that cause failure (or plastic flow) of the terrain material and cannot be used to estimate its deformation [3,4]. All these impose limitations to the applications of the theory of plastic equilibrium to the study of vehicle–terrain interaction [3].

20.4 EMPIRICAL METHODS FOR PREDICTING OFF-ROAD VEHICLE PERFORMANCE BASED ON THE CONE INDEX

In the past, owing to the lack of understanding of the mechanics of vehicle–terrain interaction, empirical methods have been developed for predicting off-road vehicle performance. Some of these methods are still being used today, primarily for operational planning or for preliminary evaluation of vehicle candidates for procurement purposes.

The general approach to the development of empirical methods for predicting off-road vehicle performance is to conduct tests of a select group of vehicles considered representative over a range of terrains of interest. The terrain is usually identified (or classified) by simple measurements or field observations. The results of vehicle performance testing and the terrain characteristics identified are then empirically correlated. This can lead to the development of empirical relationships for evaluating terrain trafficability on the one hand and vehicle mobility on the other.

Empirical methods based on the measurements made by a cone penetrometer shown in Figure 20.9 to identify terrain were originally developed by the U.S. Army Waterways Experiment Station (WES) (now known as the U.S. Army Engineer Research and Development Center, Vicksburg, Mississippi, United States) during World War II. The commonly used cone penetrometer has a 30° circular cone with a 0.5 in.^{2} (3.23 cm^2) base area.

These empirical methods were originally intended to provide military intelligence and reconnaissance personnel with a simple means to assess vehicle mobility on a "go/no-go" basis in fine- and coarse-grained soils. Fine-grained soils are silt or clayey soils for which 50% or more by weight of grains are smaller than 0.074 mm in diameter (or able to pass through a No. 200 sieve). Coarse-grained soils are beach or desert soils, usually containing less than 7% of the grains smaller than 0.074 mm in diameter, or soils containing 7% or more of the grains smaller than 0.074 mm in diameter but not in a wet condition, that is, they are non-remoldable. These empirical methods form the basis for the subsequent development of the NATO (North Atlantic Treaty Organization) Reference Mobility Model (NRMM).

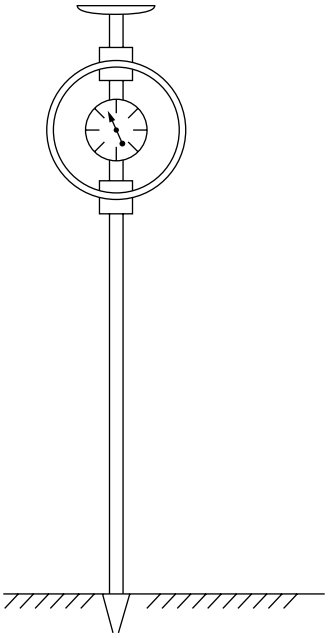


FIGURE 20.9 A cone penetrometer. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

20.4.1 Empirical Method for Predicting Tracked Vehicle Performance

In the method developed by the U.S. Army WES, an empirical equation is first used to calculate the mobility index (*MI*) of a tracked vehicle [3,6]. The *MI* for tracked vehicles is expressed by

$$\text{Mobility index} = \left(\frac{\text{contact pressure factor} \times \text{weight factor}}{\text{track factor} \times \text{grouser factor}} + \text{bogie factor} - \text{clearance factor} \right) \times \text{engine factor} \times \text{transmission factor} \tag{20.11}$$

where

$$\text{Contact pressure factor} = \frac{\text{gross weight, lb}}{\text{area of tracks in contact with ground, in.}^2}$$

Weight factor: less than 50,000 lb (222.4 kN) = 1.0
50,000 – 69,999 lb (222.4 – 311.4 kN) = 1.2
70,000 – 99,999 lb (311.4 – 444.8 kN) = 1.4
100,000 lb (444.8 kN) or greater = 1.8

Track factor = track width, in./100

Grouser factor: grouser less than 1.5 in. (3.8 cm) high = 1.0
grouser more than 1.5 in. (3.8 cm) high = 1.1

$$\text{Bogie factor} = \frac{\text{gross weight, lb, divided by 10}}{\text{total number of bogies} \times \text{area of one track shoe, in.}^2}$$

Clearance factor = clearance, in./10
Engine factor: ≥10 hp/ton of vehicle weight = 1.0
<10 hp/ton of vehicle weight = 1.05
Transmission factor: automatic = 1.0; manual = 1.05

Based on the *MI,* a parameter called the vehicle cone index (*VCI*) is calculated. The *VCI* represents the minimum strength of a soil in the critical layer that permits a given vehicle to successfully make a specific number of passes, usually one pass or 50 passes. For instance, the values of *VCI* for one pass and 50 passes, VCI_1 and VCI_{50}, for fine-grained soils may be calculated from the *MI* using the following empirical equations:

$$VCI_1 = 7.0 + 0.2MI - \left(\frac{39.2}{MI + 5.6} \right) \tag{20.12}$$

$$VCI_{50} = 19.27 + 0.43MI - \left(\frac{125.79}{MI + 7.08} \right) \tag{20.13}$$

The soil strength is described in terms of either the rating cone index (*RCI*) for fine-grained soils or the cone index (*CI*) for coarse-grained soils. The *CI* is the force per unit cone base area measured by the cone penetrometer. While the *CI* is commonly used as a parameter with no dimensions attached, it is in fact the penetration force in pounds divided by the area of the cone base in square inches, and thus, it has the unit of pressure. The *RCI* is the *CI* of the soil after being remolded following a recommended procedure [6]. The critical layer referred to earlier varies with the type and weight of the vehicle and soil strength profile. For freely draining or clean sand, it is usually the 0–6 in. (0–15 cm) layer. For fine-grained soils and poorly drained sands with fines, the critical layer is usually the 0–6 in. (0–15 cm) layer for one pass and the 6–12 in. (15–30 cm) layer for 50 passes.

After the *VCI* and soil strength have been determined, the values of the performance parameters of a tracked vehicle, such as the net maximum drawbar pull coefficient (the ratio of maximum drawbar pull to vehicle weight), maximum slope negotiable, and towed motion resistance coefficient (the ratio of towed motion resistance to vehicle weight), are then empirically determined as functions of vehicle type, number of passes to be completed, and the excess of *RCI* over *VCI* (i.e., *RCI* − *VCI*) for fine-grained soils or *CI* for coarse-grained soils [6]. For instance, Figure 20.10 shows the empirical relations between the net maximum drawbar pull coefficient (or maximum slope negotiable) and *RCI* − *VCI* for tracked vehicles with different grouser heights and for wheeled vehicles over

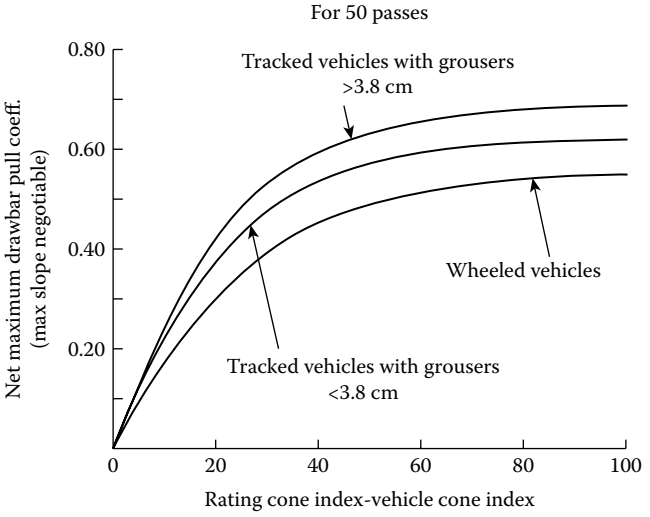


FIGURE 20.10 Variations of net maximum drawbar pull coefficient with the excess of *RCI* over *VCI* on level, fine-grained soil. (Reproduced from Rula, A.A. and Nuttall, C.J., An analysis of ground mobility models (ANAMOB), Technical Report No. M-71-4, U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, 1971.)

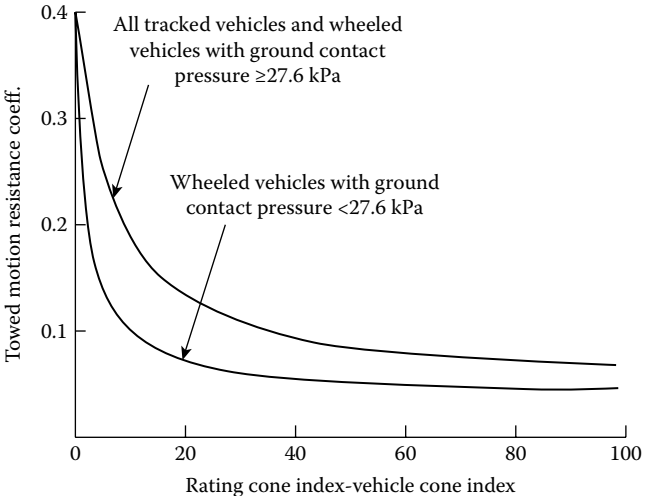


FIGURE 20.11 Variations of the first-pass towed motion resistance with the excess of *RCI* over *VCI* on level, fine-grained soil. (Reproduced from Rula, A.A. and Nuttall, C.J., An analysis of ground mobility models (ANAMOB), Technical Report No. M-71-4, U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, 1971.)

fine-grained soils. The empirical relations between the first-pass towed motion resistance coefficient and the excess of *RCI* over *VCI* for tracked and wheeled vehicles operating on fine-grained soils are shown in Figure 20.11.

20.4.2 Empirical Method for Predicting Wheeled Vehicle Performance

Similar to the empirical method developed by WES for predicting tracked vehicle performance, in this method an empirical equation is first used to calculate the *MI* of an off-road wheeled vehicle [3,6]. The *MI* for wheeled vehicles is expressed by

$$\text{Mobility index} = \left(\frac{\text{contact pressure factor} \times \text{weight factor}}{\text{tire factor} \times \text{grouser factor}} + \text{wheel load factor} - \text{clearance factor} \right) \times \text{engine factor} \times \text{transmission factor} \tag{20.14}$$

where

$$\text{Contact pressure factor} = \frac{\text{gross weight, lb}}{\text{tire width} \times \text{tire radius} \times \text{no. of tires, in.}^2}$$

Weight Factor	Weight Range (lb)	Weight Factor (Y)
	<2,000 (8.9 kN)	$Y = 0.553X$
	2,000–13,500 (8.9–60 kN)	$Y = 0.033X + 1.050$
	13,501–20,000 (60–88.9 kN)	$Y = 0.142X - 0.420$
	>20,000 (88.9 kN)	$Y = 0.278X - 3.115$

where X = gross weight, kips/no. of axles

Tire factor = (10 + tire width, in.)/100
Grouser factor: with chains = 1.05
without chains = 1.0
Wheel load factor = gross weight, kips/(no. of axles × 2)
Clearance factor = clearance, in./10
Engine factor: ≥ 10 hp/ton of vehicle weight = 1.0
< 10 hp/ton of vehicle weight = 1.05
Transmission factor: automatic = 1.0; manual = 1.05

Similar to the empirical method for predicting tracked vehicle performance, the *MI* of wheeled vehicles is used to determine a *VCI*. For a self-propelled, wheeled vehicle, the *VCI* for fine-grained soils is related to the *MI* by the following empirical equations:

For one pass, if $MI \leq 115$,

$$VCI_1 = 11.48 + 0.2MI - \left(\frac{39.2}{MI + 3.74}\right) \tag{20.15}$$

and if $MI > 115$,

$$VCI_1 = 4.1MI^{0.446} \tag{20.16}$$

For 50 passes,

$$VCI_{50} = 28.23 + 0.43MI - \left(\frac{92.67}{MI + 3.67}\right) \tag{20.17}$$

Lately, a correction factor (*CF*), taking into account the effect of tire deflection, has been introduced into the calculations of *VCI*. The *CF* is expressed by

$$CF = \left(\frac{0.15}{\delta/h}\right)^{0.25} \tag{20.18}$$

where
δ is the tire deflection
h is the unloaded section height of the tire

The corrected value of *VCI* is obtained by multiplying Equation 20.15, 20.16, or 20.17 by the *CF* given by Equation 20.18.

After the *VCI* of a wheeled vehicle and the strength of the soil have been determined, the net maximum drawbar pull coefficient (maximum slope negotiable) and the towed motion resistance coefficient of the wheeled vehicle can then be predicted using empirical relations shown in Figures 20.10 and 20.11, respectively. The values of the performance parameters are related to the excess of the *RCI* over the *VCI* for fine-grained soils.

20.5 EMPIRICAL METHODS FOR PREDICTING VEHICLE MOBILITY BASED ON THE MEAN MAXIMUM PRESSURE

Another type of empirical method for predicting the mobility of off-road vehicles is based on the concept of mean maximum pressure (*MMP*) proposed by Rowland [7,8]. *MMP* is defined as the mean value of the maxima occurring under all of the road wheel stations.

Empirical equations for predicting the value of *MMP* of track systems with different design features are given in the succeeding text.

For link and belt tracks on rigid road wheels,

$$MMP = \frac{1.26W}{2n_r A_l b\sqrt{t_t D}}, \text{kPa} \tag{20.19}$$

For belt tracks on pneumatic tired road wheels,

$$MMP = \frac{0.5W}{2n_r b\sqrt{D\delta}}, \text{kPa} \tag{20.20}$$

where

W is the vehicle weight in kN
n_r is the number of wheel stations in one track
A_l is the rigid area of link (or belt track cleat) as a proportion of $b \times t_t$
b is the width of the track or pneumatic tire in m
t_t is the track pitch in m
D is the outer diameter of the road wheel or pneumatic tire in m
δ is the radial deflection of the pneumatic tire under load in m

Table 20.1 shows the values of *MMP* for various tracked vehicles calculated using the empirical formulas described earlier [8].

To evaluate whether a particular vehicle with a specific value of *MMP* will have adequate mobility over a given terrain, a set of desirable values of the *MMP* for different levels of mobility on various types of terrain is proposed by Rowland, as shown in Table 20.2.

It should be pointed out that in the empirical Equations 20.19 and 20.20, terrain characteristics are not explicitly taken into account in the calculations of *MMP*. Thus, the values of *MMP* calculated are independent of terrain conditions. In reality, the pressure distribution under a tracked vehicle, and hence the actual value of *MMP*, is strongly influenced by terrain characteristics [3,4,9].

TABLE 20.1
Values of the Mean Maximum Pressure (*MMP*) of Some Tracked Vehicles

Vehicle	Track Configuration	Weight (kN)	Mean Maximum Pressure (kPa)
All Terrain Vehicle Bv206	Belt track	62	34
Armored Personnel Carrier M113	Link track	114	121
Caterpillar D4 tractor	Link track	59	82
Caterpillar D7 tractor	Link track	131	80
Main Battle Tank AMX 30	Link track	370	249
Main Battle Tank Challenger	Link track	608	282
Main Battle Tank Leopard I	Link track	436	235
Main Battle Tank Leopard II	Link track	542	222
Main Battle Tank M1A1	Link track	548	236
Main Battle Tank Merkava	Link track	589	279
Main Battle Tank T62	Link track	373	256
Swedish S-Tank	Line track	363	290

Source: Larminie, J.C., Modifications to the mean maximum pressure system, *Journal of Terramechanics*, *29*(2), 239–255, 1992; and Refs. [7,8].

TABLE 20.2
Desired Values of the *MMP* for Various Terrains

	MMP (kPa)		
Terrain	**Ideal (Multi-Pass Operation or Good Gradeability)**	**Satisfactory**	**Maximum Acceptable (Mostly Trafficable at Single-Pass Level)**
Wet, fine-grained			
Temperate	150	200	300
Tropical	90	140	240
Muskeg	30	50	60
Muskeg floating mat and European bogs	5	10	15
Snow	10	25–30	40

Source: Rowland, D., A review of vehicle design for soft ground operation, *Proceedings of the 5th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. 1, Detroit, MI, 1975.

It has been shown that the values of *MMP* calculated using Rowland's empirical formulas are significantly different from those measured and that the effects of vehicle design parameters on the values of *MMP* are not accurately represented by the empirical formulas on many types of terrain [9]. Furthermore, Equation 20.19 or 20.20, together with Table 20.2, can only be employed to predict vehicle performance qualitatively on a "go/no-go" basis and cannot be used to quantitatively predict the performance of the vehicle, such as its motion resistance, thrust, drawbar pull, and tractive efficiency.

In summary, empirical methods are simple to use and may be useful in estimating the performance of vehicles with design features similar to those that have been tested under similar operating conditions. It should be pointed out, however, that empirical relations should not normally be extrapolated beyond the conditions upon which they were derived. Consequently, empirical methods have inherent limitations. For instance, it is uncertain that they can play a useful role in the evaluation of vehicles with new design features or in new operating environments.

20.6 SEMIEMPIRICAL METHODS FOR PREDICTING THE PERFORMANCE OF TRACKED VEHICLES

20.6.1 Bevameter Techniques

One of the widely known semiempirical methods for predicting the performance of off-road vehicles is that due to M.G. Bekker [2–4]. This method is based on terrain data measured using an instrument conceived by Bekker, which has now become known as the bevameter (Bekker value meter). The bevameter was developed on the premise that terrain properties should be measured under loading conditions similar to that of the vehicle running gear. The vehicle running gear exerts normal load to the terrain surface resulting in sinkage and causing motion resistance. It also applies shear load to the terrain surface resulting in the development of thrust and the associated slip, as shown in Figure 20.12. To simulate these, the bevameter is designed to perform two sets of tests: one is a set of penetration tests and the other is a set of shear tests. In the penetration tests, a plate of suitable size is used to simulate the contact area of the vehicle running gear, and the pressure–sinkage (p–z) relationship of the terrain is measured, as shown in Figure 20.12. This relationship is used to predict the normal pressure distribution on the vehicle–terrain interface, as well as vehicle sinkage. To minimize the uncertainty

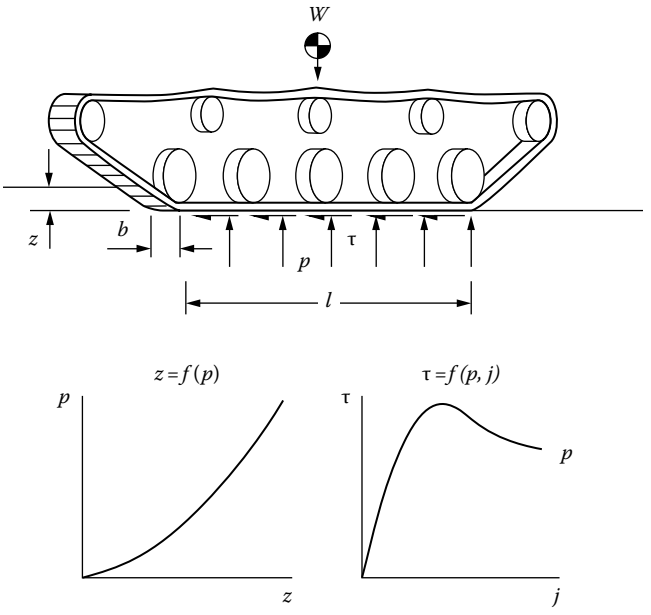


FIGURE 20.12 The interaction between a vehicle running gear and terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

in applying the measured data to the prediction of the performance of full-size vehicles, it is preferable that the size of the plate used in the penetration tests be comparable to that of the contact area of a track link or a tire. In the shear tests, a shear ring or a rectangular shear plate is used to simulate the shearing action of the vehicle running gear. The shear stress–shear displacement relationship (τ–j), as shown in Figure 20.12, and the shear strength of the terrain (as described in Section 20.3) are measured. This provides the necessary data for predicting the shear stress distribution on the vehicle–terrain interface and the thrust–slip relationship of the vehicle [2–4].

After terrain data have been collected by the bevameter, they should be properly processed and characterized so that they may be integrated into a framework for predicting the performance of off-road vehicles [3,4].

20.6.1.1 Pressure–Sinkage Relationships

A set of pressure–sinkage curves measured in a sandy loam is shown in Figure 20.13. For most mineral terrain, Bekker proposed the following equation for characterizing the pressure–sinkage relationship:

$$p = \left(\frac{k_c}{b} + k_\phi \right) z^n \tag{20.21}$$

where

p is the pressure
b is the characteristic dimension of the plate, for a rectangular plate it is the width and for a circular plate it is the radius
z is the sinkage
n, k_c, and k_ϕ are pressure–sinkage parameters

For organic terrain (muskeg), snow-covered terrain with ice layers, and others, the pressure–sinkage relationships are different from that for mineral terrain described by Equation 20.21.

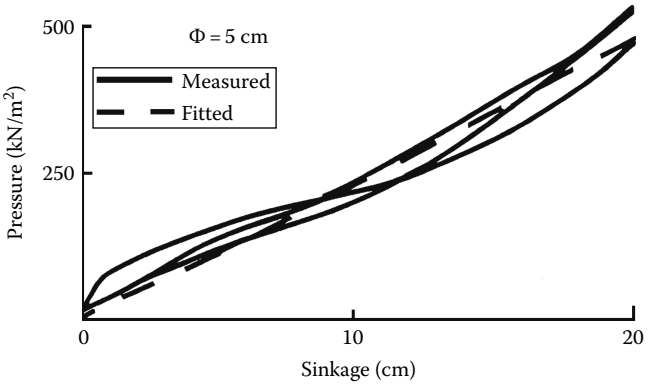


FIGURE 20.13 Pressure–sinkage curves for a sandy loam. (Reproduced from Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.)

TABLE 20.3
Terrain Values

Terrain	Moisture Content (%)	n	k_c (kN/m^{n+1})	k_ϕ (kN/m^{n+2})	c (kPa)	ϕ (Degrees)
Dry sand (Land Locomotion Lab., LLL)	0	1.1	0.99	1528.43	1.04	28
LETE sand (Wong)		0.79	102	5301	1.3	31.1
Sandy loam (LLL)	15	0.7	5.27	1515.04	1.72	29
Rubicon sandy loam (Wong)	43	0.66	6.9	752	3.7	29.8
Upland sandy loam (Wong)	51	1.10	74.6	2080	3.3	33.7
Clayey soil (Thailand)	38	0.5	13.19	692.15	4.14	13
Heavy clay (WES)	40	0.11	1.84	103.27	20.69	6
Lean clay (WES)	32	0.15	1.52	119.61	13.79	11
North Gower clayey loam (Wong)	46	0.73	41.6	2471	6.1	26.6
Grenville loam (Wong)	24	1.01	0.06	5880	3.1	29.8
Snow (United States)		1.6	4.37	196.72	1.03	19.7
Snow (Sweden)		1.44	10.55	66.08	6	20.7

Sources: Bekker, M.G., *Introduction to Terrain-Vehicle Systems*, University of Michigan Press, Ann Arbor, MI, 1969; Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008; and Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.

Detailed analysis of the pressure–sinkage characteristics of organic terrain (muskeg) and snow-covered terrain with ice layers may be found in Refs. [3,4,10,11]. The procedure for deriving the pressure–sinkage parameters from test data is described in Refs. [2–4]. Table 20.3 gives the values of n, k_c, and k_ϕ of a sample of terrains. More terrain data may be found in Refs. [2–4,10,12].

20.6.1.2 Shear Stress–Shear Displacement Relationships

A shear curve measured in a mineral terrain is shown in Figure 20.14. This type of shear curve can be characterized by an exponential equation of the following form:

$$\tau = \tau_{max}(1 - e^{-j/k}) = (c + \sigma \tan \phi)(1 - e^{-j/k}) \tag{20.22}$$

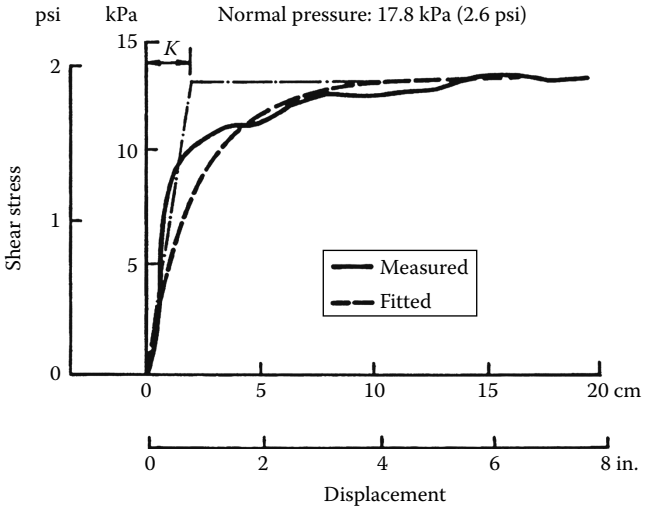


FIGURE 20.14 A shear curve of a simple exponential form. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

where

τ is the shear stress

τ_{max} is the maximum shear stress or shear strength of the terrain

c and ø are the cohesion and angle of internal shearing resistance of the terrain, respectively

j is the shear displacement

K is referred to as the shear deformation modulus

K may be considered as a measure of the magnitude of shear displacement required to develop the maximum shear stress. The value of K determines the shape of the shear curve. Its value may be represented by the distance between the vertical axis and the point of intersection of the straight line tangent to the shear curve at the origin and the horizontal line representing the maximum shear stress, as shown in Figure 20.14. The slope of the shear curve at the origin can be obtained by differentiating τ with respect to j in Equation 20.22:

$$\frac{d\tau}{dj} = \frac{\tau_{max}}{K} e^{-j/k}; \quad \text{when} \quad j = 0, \quad \frac{d\tau}{dj} = \frac{\tau_{max}}{K} \tag{20.23}$$

Thus, the value of K can be determined from the slope of the shear curve at the origin and the maximum shear stress τ_{max}. The value of K may also be taken as 1/3 of the shear displacement where the shear stress τ is 95% of the maximum shear stress τ_{max}.

For organic terrain (muskeg), brittle soil, frozen snow, etc., the shear stress–shear displacement relationships usually cannot be characterized by Equation 20.22. Detailed analysis of the shearing characteristics of organic terrain (muskeg), brittle soil, and frozen snow may be found in Refs. [3,4,13].

20.6.1.3 Response of Terrain to Repetitive Loading

When an off-road vehicle is in a straight line motion, an element of the terrain under the running gear is first subject to the load applied by the front wheel (or the leading road wheel in a track system). When the front wheel has passed, the load on the terrain element is released. Load is reapplied as the succeeding wheel rolls over it. A terrain element is thus subject to the repetitive loading of the consecutive wheels of a multi-axle wheeled vehicle or the road wheels of a tracked vehicle.

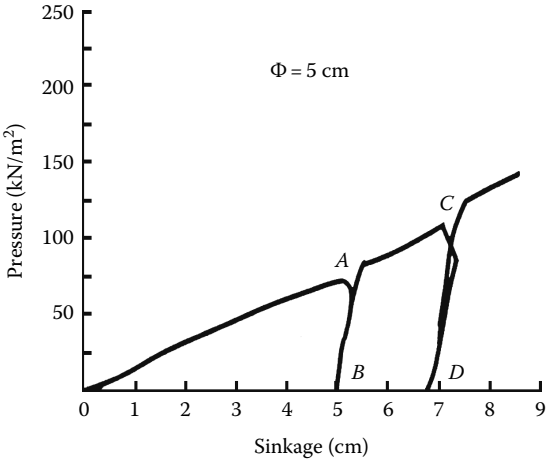


FIGURE 20.15 Response to repetitive normal load of a mineral terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

The loading–unloading–reloading cycle continues until the rear wheel of the vehicle has passed over it. To realistically predict the normal pressure distribution under a moving vehicle and hence its performance, the response of the terrain to repetitive loading must be measured. The response of a mineral terrain to repetitive normal load is shown in Figure 20.15. It can be seen that the pressure initially increases with sinkage along *OA*. However, when the load applied to the terrain is reduced at *A*, the pressure–sinkage relationship during unloading follows line *AB*. When the load is reapplied at *B,* the pressure–sinkage relationship follows, more or less, the same path as that during unloading. When the reapplied load exceeds that at which the preceding unloading begins (point *A* in Figure 20.15), additional sinkage results. With the further increase of pressure, the pressure–sinkage relationship appears to follow the continuous loading path as *AC* shown in the figure. The characteristics of the second unloading–reloading cycle that begins at *C* are similar to those of the first one. For organic terrain (muskeg) or snow-covered terrain, the response to repetitive loading is different from that shown in Figure 20.15. Further information may be found in Refs. [3,4].

It should be emphasized that to realistically predict the performance of off-road vehicles, the response of terrain to repetitive loading must be taken into consideration, in addition to the pressure–sinkage and shearing characteristics described in previous sections.

20.6.2 Motion Resistance of a Track

In the semiempirical method proposed by Bekker, it is assumed that a track is equivalent to a rigid footing and that the normal reaction exerted on the track by the terrain can be equated to that beneath a sinkage plate at the same depth in a pressure–sinkage test, as described in Section 20.6.1.1. If the center of gravity of the vehicle is located at the midpoint of the track contact length, the normal pressure distribution may be assumed to be uniform. Using the pressure–sinkage relationship given by Equation 20.21, the sinkage z_0 of the track is expressed by

$$z_0 = \left(\frac{p}{k_c/b + k_\phi} \right)^{1/n} = \left(\frac{W/bl}{k_c/b + k_\phi} \right)^{1/n} \tag{20.24}$$

where

p is the normal pressure

W is the normal load on a track

b and l are the width and length of the track in contact with the terrain, respectively

The work done in compacting the terrain and making a rut of width b, length l, and depth z_0 is given by

$$\text{Work} = bl\int_0^{z_o} pdz = bl\int_0^{z_o}\left(\frac{k_c}{b}+k_\phi\right) z^n\, dz = bl\left(\frac{k_c}{b}+k_\phi\right)\left(\frac{z_0{}^{n+1}}{n+1}\right) \tag{20.25}$$

Substituting for z_0 from Equation 20.24 yields

$$\text{Work} = \frac{bl}{(n+1)(k_c/b+k_\phi)^{1/n}}\left(\frac{W}{bl}\right)^{(n+1)/n} \tag{20.26}$$

If the track is pulled a distance l in the horizontal direction, the work done by the pulling force, which is equal to the magnitude of the motion resistance due to terrain compaction R_c, can be equated to the vertical work done in making a rut of length l, as expressed by Equation 20.26:

$$R_c l = \frac{bl}{(n+1)(k_c/b+k_\phi)^{1/n}}\left(\frac{W}{bl}\right)^{(n+1)/n} \tag{20.27}$$

and

$$R_c = \frac{1}{(n+1)b^{1/n}(k_c/b+k_\phi)^{1/n}}\left(\frac{W}{l}\right)^{(n+1)/n} \tag{20.28}$$

Equation 20.28 is for predicting the motion resistance due to compacting the terrain by a track with uniform pressure distribution based on Bekker's pressure–sinkage equation (Equation 20.21). Expressions for motion resistance based on other pressure–sinkage relations may be derived in a similar way.

As the semiempirical method proposed by Bekker is based on a number of simplifying assumptions, it can only provide a preliminary assessment of the motion resistance of a track. For instance, the idealization of a track as a rigid footing is not realistic, particularly for tracked vehicles designed for high-speed operations, such as military vehicles with relatively short track pitch. For this type of tracked vehicle, the normal pressure distribution under the track is far from uniform [3,4]. To improve the accuracy of prediction, more realistic and comprehensive methods have been developed and will be discussed in the Chapter 34.

20.6.3 Thrust–Slip Relationship of a Track

While Equation 20.9 can be used to estimate the maximum thrust developed by a track at high slips (i.e., the track spins without the corresponding forward movement), it is desirable to determine the relationship between the thrust and slip over the full operating range. This is because tracked vehicles usually operate at relatively low slips in order to achieve high operating efficiency. To predict the relation between thrust and slip, it is necessary to examine the development of shear displacement beneath the track, since shear stress is a function of shear displacement, as described in Section 20.6.1.2. The shear displacement at various points under the track is shown schematically in Figure 20.16. At point 1 at the front of the track, the grouser just comes into contact with the terrain. It cannot develop the same shear displacement as the other grousers behind it, since they have been shearing the terrain surface for varying periods of time. The amount of shear displacement j

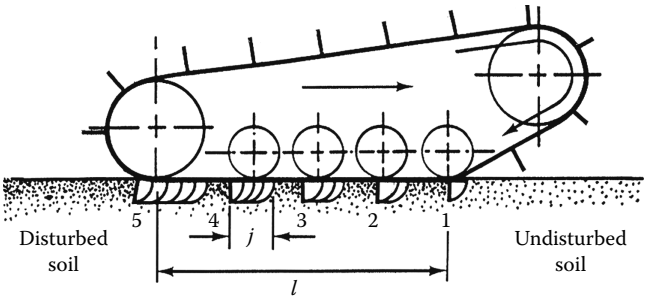


FIGURE 20.16 Development of shear displacement under a track. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956, with permission of the University of Michigan Press.)

increases along the track contact length and reaches its maximum value at the rear of the contact area. To examine the development of shear displacement beneath a track quantitatively, the slip i is defined first [1,3]:

$$i = 1 - \frac{V}{r\omega} = 1 - \frac{V}{V_t} = \frac{V_t - V}{V_t} = \frac{V_j}{V_t} \tag{20.29}$$

where

V is the actual forward speed of the track

V_t is the theoretical speed of the track that is equa1 to the product of the angular speed ω and the pitch radius r of the sprocket

V_j is the speed of slip of the track with respect to the terrain

When the track is slipping, V_j will be in the direction opposite to that of vehicle motion. Since the track is assumed to be inextensible, the speed of slip V_j is the same for every point on the track in contact with the terrain. The shear displacement j at the point located at a distance x from the front of the contact length can be determined by

$$j = V_j t \tag{20.30}$$

where t is the contact time of the point in question and is equal to x/V_t.

Rearranging Equation 20.30, the expression for shear displacement j becomes

$$j = \frac{V_j x}{V_t} = ix \tag{20.31}$$

This indicates that the shear displacement beneath a flat track increases linearly from the front to the rear of the contact length. Since the shear stress is related to shear displacement as discussed in Section 20.6.1.2, the shear stress distribution on the track contact area can be found. For a flat track with uniform normal pressure and with the shear stress–shear displacement relationship described by Equation 20.22, the thrust–slip relationship for a track is given by

$$F = b\int_0^l \left(c + \frac{W}{bl}\tan\phi\right)(1 - e^{-ix/K})dx = (Ac + W\tan\phi)\left[1 - \frac{K}{il}(1 - e^{-il/k})\right] \tag{20.32}$$

Equation 20.32 gives the functional relationship among thrust F; vehicle design parameters A, b, l, and W; terrain parameters, c, ø, and K; and track slip i. If the slip is 100%, Equations 20.9 and 20.32 are practically identical.

It should be pointed out that the thrust–slip relationship for a track given by Equation 20.32 is based on a number of simplifying assumptions. In practice, the track usually does not behave like a rigid footing and the normal pressure is not uniformly distributed over the contact area. Consequently, Equation 20.32 gives approximate results only [3,4]. Improved methods for predicting thrust–slip relationship of tracked vehicles will be discussed in the Chapter 34.

20.7 SEMIEMPIRICAL METHODS FOR PREDICTING THE PERFORMANCE OF WHEELED VEHICLES

20.7.1 Motion Resistance of a Rigid Wheel

While pneumatic tires have been widely used in off-road wheeled vehicles, the mechanics of rigid wheels operating over unprepared terrain is still of interest as a pneumatic tire at high inflation pressure may behave like a rigid wheel in soft soil. Furthermore, rigid wheels are still used in special circumstances, such as in paddy field or as traction-aided devices for pneumatic tires on wet terrain. One of the earlier methods for predicting the motion resistance of a rigid wheel is that proposed by Bekker [1,3]. In developing this method, it is assumed that the terrain reaction at all points on the contact area is purely radial and is equa1 to the normal pressure beneath a horizontal plate in a pressure–sinkage test described in Section 20.6.1.1 (Figure 20.17). The shear stress on the wheel–terrain interface is neglected. In the absence of shear stress, the equilibrium equations of a rigid wheel in the horizontal and vertical directions can be written as

$$R_c = b\int_0^{\theta_0} \sigma r \sin\theta \, d\theta \tag{20.33}$$

$$W = b\int_0^{\theta_0} \sigma r \cos\theta \, d\theta \tag{20.34}$$

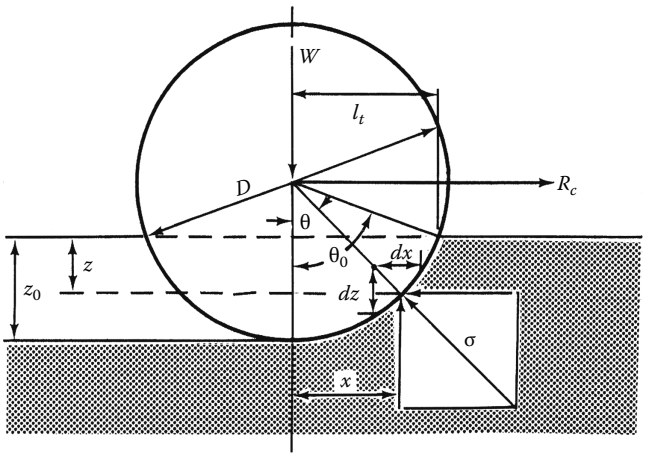


FIGURE 20.17 A simplified wheel–soil interaction model. (Reproduced from Bekker, M.G., *Theory of Land Locomotion*, University of Michigan Press, Ann Arbor, MI, 1956, with permission of the University of Michigan Press.)

where
R_c is the motion resistance
W is the load on the wheel
σ is the normal pressure exerted on the wheel rim by the terrain
b and r are width and radius of the wheel, respectively

Since it is assumed that the normal pressure σ acting on the wheel rim is equal to the normal pressure p beneath a plate at the same depth z in a pressure–sinkage test, then $\sigma\, r \sin\theta\, d\theta = p\, dz$ and $\sigma\, r \cos\theta\, d\theta = p\, dx$. Using the pressure–sinkage relationship defined by Equation 20.21, Equation 20.33 becomes

$$R_c = b\int_0^{Z_0}\left(\frac{k_c}{b}+k_\phi\right)z^n dz = b\left[\left(\frac{k_c}{b}+k_\phi\right)\frac{z_0^{n+1}}{n+1}\right] \tag{20.35}$$

The value of R_c determined by Equation 20.35 is equivalent to the vertical work done per unit length in pressing a plate of width b into the terrain to a depth of z_0. The assumption for the stress distribution made by Bekker implies that the motion resistance of a rigid wheel is due to the vertical work done in making a rut of depth z_0. The rut depth z_0 can be derived by considering the equilibrium of the wheel in the vertical direction using Equation 20.34 and is expressed by [1,3]

$$z_0 = \left[\frac{3W}{b(3-n)(k_c/b+k_\phi)\sqrt{D}}\right]^{2/(2n+1)} \tag{20.36}$$

Substituting Equation 20.36 into Equation 20.35, the compaction resistance R_c is expressed by

$$R_c = \left[\frac{1}{(3-n)^{(2n+2)/(2n+1)}(n+1)b^{1/(2n+1)}(k_c/b+k_\phi)^{1/(2n+1)}}\right]\left[\frac{3W}{\sqrt{D}}\right]^{(2n+2)/(2n+1)} \tag{20.37}$$

It can be seen from Equation 20.37 that to reduce the motion resistance, it seems more effective to increase the wheel diameter D than the width b, as D enters the equation in higher power than b.

20.7.2 Motion Resistance of a Pneumatic Tire

The motion resistance of a pneumatic tire depends on its mode of operation. If the ground is sufficiently soft and the sum of the inflation pressure p_i and the pressure produced by the stiffness of the carcass p_c is greater than the maximum pressure that the terrain can support at the lowest point of the tire circumference, the tire will remain round like a rigid rim. This is referred to as the rigid mode of operation. On the other hand, if the terrain is firm and the tire inflation pressure is low, a portion of the circumference of the tire will be flattened. This is referred to as the elastic mode of operation. To predict the motion resistance of a tire, it is necessary, first of all, to determine the operating mode of the tire. If the tire behaves like a rigid rim, using Equation 20.21, the normal pressure p_g at the lowest point of contact at sinkage z_0 (i.e., the bottom dead center) is given by

$$p_g = \left(\frac{k_c}{b}+k_\phi\right)z_0^n \tag{20.38}$$

Substituting Equation 20.36 into Equation 20.38, the expression for p_g becomes

$$p_g = \left(\frac{k_c}{b} + k_\phi\right)^{1/(2n+1)} \left[\frac{3W}{(3-n)b\sqrt{D}}\right]^{2n/(2n+1)} \tag{20.39}$$

If the sum of the inflation pressure p_i and the pressure due to carcass stiffness p_c is greater than the pressure defined by Equation 20.39, which may be called the critical pressure p_{gcr}, the tire will remain round like a rigid wheel [3]. Under this condition, the motion resistance due to compacting the terrain can be predicted using Equation 20.37. On the other hand, if it is less than the critical pressure p_{gcr}, the tire is in the elastic mode of operation and a portion of the tire circumference will be flattened. The contact pressure on the flattened portion will be equa1 to $p_i + p_c$. In this case, the sinkage of the tire z_0 can be predicted using Equation 20.21:

$$z_0 = \left[\frac{p_i + p_c}{k_c/b + k_\phi}\right]^{1/n} \tag{20.40}$$

Substituting Equation 20.40 into Equation 20.35, the motion resistance of a tire in the elastic mode of operation is given by

$$R_c = b\left(\frac{k_c}{b} + k_\phi\right)\left(\frac{z_0^{n+1}}{n+1}\right) = \frac{b(p_i + p_c)^{(n+1)/n}}{(n+1)(k_c/b + k_\phi)^{1/n}} \tag{20.41}$$

Recent studies show that the sum of the inflation pressure and the pressure produced by carcass stiffness, $p_i + p_c$, can be represented by the average ground pressure p_{gr} of the tire on hard ground [2–4]. The average ground pressure p_{gr} may be calculated from the normal load and the gross contact area of the tire on hard surface given in the generalized deflection chart of the tire provided by the tire manufacturer [2–4].

20.7.3 Thrust–Slip Relationship of a Wheel

In the wheel–terrain interaction model proposed by Bekker and described in Section 20.7.1, only the normal pressure on the interface is considered and the shear stress is completely neglected. As a result, the semiempirical method developed by Bekker is unable to predict the tractive performance of a driven wheel. Later, a method was developed by Wong and Reece to predict the shear stress distribution on the wheel–terrain interface and the thrust–slip relationship of a wheel [3,4,14].

To predict the thrust–slip relationship of a wheel, the development of shear displacement along the wheel–terrain interface has to be examined first. The shear displacement developed along the contact area of a wheel may be determined by analyzing the slip velocity V_j on the wheel–terrain interface. For a rigid wheel, the slip velocity V_j of a point on the rim relative to the terrain is the tangential component of the absolute velocity at that point, as illustrated in Figure 20.18. The magnitude of the slip velocity V_j of a point on the rim defined by angle θ can be expressed by

$$V_j = r\omega[1-(1-i)\cos\theta] \tag{20.42}$$

The shear displacement j at a point defined by angle θ on the wheel–terrain interface is given by

$$j = \int_0^t V_j dt = \int_\theta^{\theta_0} r\omega[1-(1-i)\cos\theta]\frac{d\theta}{\omega} = r[(\theta_0 - \theta) - (1-i)(\sin\theta_0 - \sin\theta)] \tag{20.43}$$

where θ_0 is the entry angle where a point on the rim first comes into contact with the terrain (Figure 20.18).

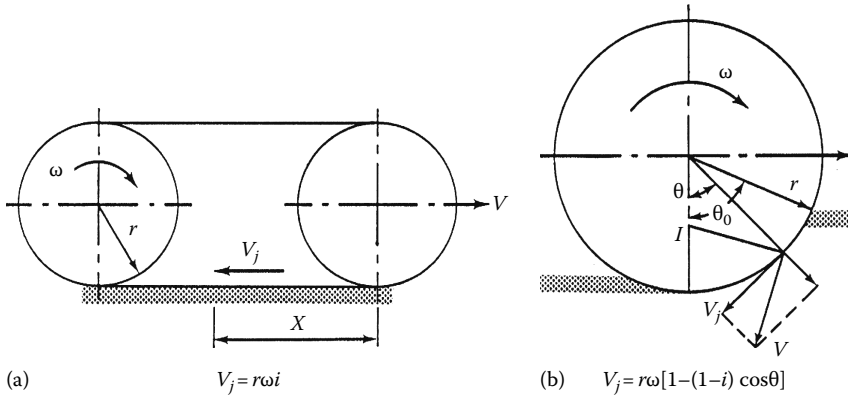


FIGURE 20.18 Slip velocity on (a) track–terrain and (b) wheel–terrain interface. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

Based on the relationship between the shear stress and shear displacement discussed previously in Section 20.6.1.2, the shear stress distribution along the contact area of a rigid wheel can be determined. For instance, using Equation 20.22, the shear stress distribution along the wheel–terrain interface is expressed by

$$\tau(\theta) = [c + p(\theta)\tan\phi](1 - e^{-j/K}) = \tau_{max}[1 - e^{-(r/K)[\theta_0 - \theta - (1-i)(\sin\theta_0 - \sin\theta)]}] \tag{20.44}$$

The normal pressure distribution along a rigid wheel $p(\theta)$ may be predicted by a variety of methods, including the simplified method proposed by Bekker and described in Section 20.7.1. By integrating the horizontal component of shear (tangential) stress over the entire contact area, the thrust F developed by the wheel is expressed by

$$F = rb\int_0^{\theta_0} \tau(\theta)\cos\theta\, d\theta \tag{20.45}$$

It should be mentioned that taking into account the presence of shear stress on the wheel–terrain interface, the vertical load on the wheel is partly supported by its vertical component. This fact has been neglected in the original wheel–terrain interaction model proposed by Bekker. In a more complete analysis of wheel–terrain interaction, the effects of shear stress on supporting vertical load and on the development of thrust should be taken into consideration. The equations for predicting the performance of a rigid wheel are given by [3,4]

For vertical load,

$$W = rb\left[\int_0^{\theta_0} p(\theta)\cos\theta\, d\theta + \int_0^{\theta_0} \tau(\theta)\sin\theta\, d\theta\right] \tag{20.46}$$

for drawbar pull,

$$F_d = rb\left[\int_0^{\theta_0} \tau(\theta)\cos\theta\, d\theta - \int_0^{\theta_0} p(\theta)\sin\theta\, d\theta\right] \tag{20.47}$$

and for driving torque,

$$M_w = r^2 b \int_0^{\theta_0} \tau(\theta)\, d\theta \tag{20.48}$$

Following a similar approach, the thrust–slip relationship and the overall performance of a tire in the elastic operating mode may be predicted [3,4].

20.8 CLOSING REMARKS

Different methods are available for evaluating different aspects of off-road vehicle–terrain interaction. The theory of elasticity may be used to predict the stress distribution in dense soil under vehicular loads. The theory of plastic equilibrium may be applied to estimating the maximum thrust that an off-road vehicle may develop, as well as the force developed by the lug of a wheel. Empirical methods may be employed to predict the performance of vehicles with design features similar to those that have been tested under similar operating conditions. However, empirical relations should not be extrapolated beyond the conditions upon which they were derived. Consequently, it is uncertain that empirical methods can play a useful role in the prediction of the performance of off-road vehicles with new design features or in new operating environment. The semiempirical methods for predicting the performance of tracked and wheeled vehicles developed by Bekker are useful in providing a basic understanding of the physical nature of off-road vehicle–terrain interaction. They also identify the vehicle and terrain parameters that would significantly affect the performance of off-road vehicles. It should be pointed out, however, that some of the assumptions used in the development of these semiempirical methods are unrealistic or oversimplified. For instance, the assumption that a track may be idealized as a rigid footing with uniform contact pressure is unrealistic in many cases. The assumption that only normal reaction (pressure) is present on the wheel–terrain interface is oversimplified. Comprehensive and realistic computer-aided methods for predicting the performance of off-road vehicles have emerged recently and will be presented in the Chapter 34 [15–32].

Notations

A	contact area
b	width
c	cohesion
D	diameter
F	thrust
F_d	drawbar pull
F_p	force on a lug or a blade
h	height
j	shear displacement
K	shear deformation modulus
k_c	pressure–sinkage parameter
k_ϕ	pressure–sinkage parameter
l	contact length
n	pressure–sinkage parameter
p	pressure
R_c	motion resistance
r	radius
W	weight, normal load

z sinkage
γ_s weight density
δ tire deflection
θ angle
τ shear stress
v concentration factor
σ normal pressure
σ_r radial stress
σ_z vertical stress
ϕ angle of shearing resistance

ACKNOWLEDGMENTS

Technical assistance provided by Dr. Wei Huang in the preparation of this chapter is greatly appreciated.

REFERENCES

1. Bekker. M.G., *Theory of Land Locomotion*. University of Michigan Press, Ann Arbor, MI, 1956.
2. Bekker, M.G., *Introduction to Terrain-Vehicle Systems*. University of Michigan Press, Ann Arbor, MI, 1969.
3. Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.
4. Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.
5. Sohne, W., Fundamentals of pressure distribution and soil compaction under tractor tires. *Agricultural Engineering*, 39(5): 276–281, May 1958.
6. Rula, A.A. and Nuttall, C.J., An analysis of ground mobility models (ANAMOB). Technical Report No. M-71-4, U.S. Army Corps of Engineers Waterways Experiment Station, Vicksburg, MS, 1971.
7. Rowland, D., Tracked vehicle ground pressure and its effect on soft ground performance. *Proceedings of the 4th International Conference of the International Society or Terrain-Vehicle Systems,* Vol. 1, Stockholm, Sweden, 1972.
8. Rowland, D., A review of vehicle design for soft ground operation. *Proceedings of the 5th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. 1, Detroit, MI, 1975.
9. Wong, J.Y., On the role of mean maximum pressure as an indicator of cross-country mobility for tracked vehicles. *Journal of Terramechanics*, 31(3): 197–213, 1994.
10. Wong, J.Y., Radforth, J.R., and Preston-Thomas, J., Some further studies of the mechanical properties of muskeg. *Journal of Terramechanics*, 19(2): 217–223, 1982.
11. Wong, J.Y. and Preston-Thomas, J., On the characterization of the pressure-sinkage relationship of snow covers containing an ice layer. *Journal of Terramechanics*, 20(1): 1–12, 1983.
12. Wong, J.Y., Evaluation of soil strength measurements. Report No. NRCC 22881, Division of Energy, National Research Council of Canada, Ottawa, Ontario, Canada, 1983.
13. Wong, J.Y. and Preston-Thomas, J., On the characterization of the shear stress-displacement relationship of terrain. *Journal of Terramechanics*, 19(4): 225–234, 1983.
14. Wong, J.Y. and Reece, A.R., Prediction of rigid wheel performance based on the analysis of soil-wheel stresses, Part I and Part II. *Journal of Terramechanics*, 4(1 and 2): 81–98, 1967.
15. Wong, J.Y. and Preston-Thomas, J., Investigation into the effects of suspension characteristics and design parameters on the performance of tracked vehicles using an advanced computer simulation model. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Transport Engineering*, 202(D3): 143–161, 1988.
16. Wong, J.Y., Optimization of the tractive performance of articulated tracked vehicles using an advanced computer simulation model. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Journal of Automobile Engineering*, 206(Dl): 29–45, 1992.
17. Wong, J.Y., Computer-aided methods for design evaluation of track systems. *Transactions of the Society of Automotive Engineers*, Section 2, *Journal of Commercial Vehicles*, Paper No. 941675, 1994.
18. Gao, Y. and Wong, J.Y., The development and validation of a computer-aided method for design evaluation of tracked vehicles with rigid links. *Proceedings of the Institution of Mechanical Engineers*, *Part D. Journal of Automobile Engineering*, 208(D3): 207–215, 1994.

19. Wong, J.Y., Application of the computer simulation model NTVPM-86 to the development of a new version of the infantry fighting vehicle ASCOD. *Journal of Terramechanics*, 32(1): 53–61, 1995.
20. Wong, J.Y., Dynamics of tracked vehicles. *Vehicle System Dynamics*, 28(2 and 3): 7–15, 1997.
21. Wong, J.Y., Optimization of design parameters of rigid-link track systems using an advanced computer-aided method. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 212(D3): 153–167, 1998.
22. Wong, J.Y., Computer-aided methods for design evaluation of tracked vehicles and their applications to product development. *International Journal of Vehicle Design*, 22(1 and 2): 73–92, 1999.
23. Liu, C.H., Wong, J.Y. and Mang, H.A., Large strain finite element analysis of sand: model, algorithm and application to numerical simulation of tire-sand interaction. *Computers and Structures*, 74(3): 253–265, 2000.
24. Wong, J.Y. and Huang, Wei, An investigation into the effects of initial track tension on soft ground mobility of tracked vehicles using an advanced computer simulation mode. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 220: 695–711, 2006.
25. Wong, J.Y. and Huang, Wei, Study of detracking risks of track systems. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 220: 1235–1253, 2006.
26. Wong, J.Y. and Huang, Wei, "Wheels vs. Tracks" – A fundamental evaluation from the traction perspective. *Journal of Terramechanics*, 43(1): 27–42, 2006.
27. Wong, J.Y. and Huang, Wei, Approaches to improving the mobility of military tracked vehicles on soft terrain. *International Journal of Heavy Vehicle Systems*, 15(2/3/4): 127–151, 2008.
28. Wong, J.Y. and Asnani, V.M., Study of the correleation between the performances of lunar vehicle wheels predicted by the Nepean wheeled vehicle performance model and test data. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 220: 1939–1954, 2008.
29. Wong, J.Y., Development of high-mobility tracked vehicles for over snow operations. *Journal of Terramechanics*, 46(4): 141–155, 2009.
30. Wong, J.Y., Computer-aided methods for performance and design evaluation of off-road vehicles from the traction perspective. *SAE International Technical Paper* 2011-01-0266, 2011.
31. Wong, J.Y., Predicting the performances of rigid rover wheels on extraterrestrial surfaces based on test results obtained on earth, *Journal of Terramechanics*, 49(1): 49–61, 2012.
32. Wong, J.Y. and Kobayashi, T., Further study of the method of approach to testing the performance of extraterrestrial rovers/rover wheels on earth. *Journal of Terramechanics*, 49(6): 349–362, 2012.

Part IV

Modeling of Vehicle Subsystems

21 Suspension Systems

Wolfgang Matschinsky

CONTENTS

21.1 Introduction 727
21.2 Analysis of Suspensions 728
21.3 Traction and Braking 735
21.4 Cornering 739
21.5 Steering Geometry 745
21.6 Springs and Suspension 753
21.7 Elasto-Kinematics 757
21.8 Conclusion 765
Notations 766
References 767

21.1 INTRODUCTION

For the first motor vehicles, rigid axles with leaf springs were adopted from horse-drawn carriages. With growing knowledge of vehicle-dynamic laws, and of the vibrational behavior of the vehicle, kinematic suspension mechanisms were introduced that enabled additional properties like, for instance, traction and braking anti-pitch arrangements. Improved measuring methods and computer programs allow the deliberate elasto-kinematic harmonization of suspensions to achieve optimum driving stability along with optimum ride comfort.

Figure 21.1 shows a steerable independent front suspension with typical elements of construction: "a" is the wheel with the tire and "b" the wheel carrier that determines the attitude of the wheel in relation to the vehicle body. The brake "c" is arranged between the wheel and the wheel carrier. The lower transverse link "d" is triangulated by a tension link "e" to the vehicle body; "e" is here the arm of a stabilizer bar or "antiroll bar" (MacPherson principle). The damper "f" is fixed to the wheel carrier, and its piston rod represents another link of the suspension. The track rod "g" connects the steering arm of the wheel carrier with the rack-and-pinion steering "h" and so completes the suspension mechanism. The spring "i" can be regarded as a "highly compliant" suspension link, together with which the suspension becomes a "statically determinate system." The driveshaft "k" connects the wheel axle with the inboard final-drive gearbox (not shown).

The kinematic laws of the mechanisms of wheel suspensions and methods for the analysis of suspension geometry as well as suspension characteristics will be covered in this treatise.

Since the wheel and the suspension links move inside the vehicle body, and since clearance is one of the most important problems during design, every wheel suspension is usually designed on the assumption that the vehicle body is the "fixed" system (Figure 21.2). In the vehicle-fixed coordinate system used in the following, the x-axis is directed forward, the y-axis to the left, and the z-axis upward. The pitch angle ϑ, the roll angle φ, and the yaw angle ψ are also defined in this illustration. The steering angle δ on the left wheel is defined as greater than zero for a left-hand corner (this means, of course, that a positive "toe-in" angle δ_v corresponds to a negative steering angle δ).

Suspension systems are mostly represented in the following without showing the spring elements, because normally any type of spring can be used in any suspension design.

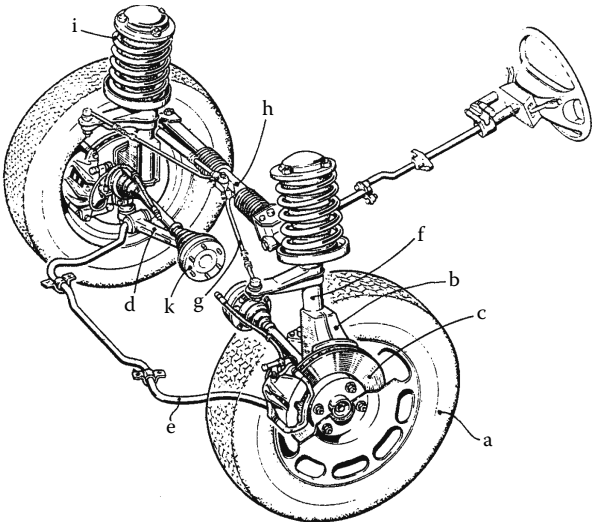


FIGURE 21.1 Steerable independent suspension with driveshafts. (Courtesy of Audi AG, Ingolstadt, Germany, 1976.)

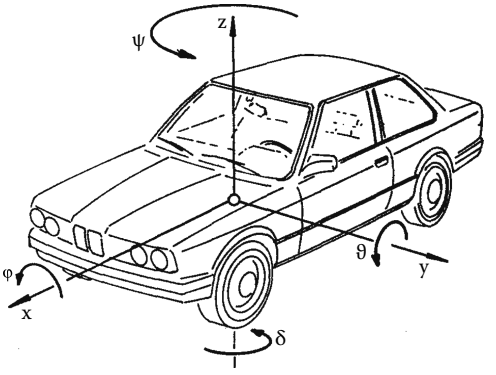


FIGURE 21.2 Coordinate system.

Suspension design is a very special topic, and the relatively small number of suspension design experts is concentrated in the companies of motorcar producers—the number of which is not very great. The special knowledge about suspensions is handed down by the more experienced to the junior designer according to the principle of "learning by doing." Theoretical treatment in publications is often simplifying, and even a book cannot mention all the special tricks applied in everyday design work. Nevertheless, an attempt will be made here to elucidate the most important considerations and methods for suspension design. This treatise closely takes pattern from the author's books on suspensions [1,6], which may be recommended for a more detailed information.

21.2 ANALYSIS OF SUSPENSIONS

To respond to road irregularities, every wheel of a fast road vehicle needs a possibility of mainly vertically directed motion, one "degree of freedom," the wheel's path following a defined and reproducible function.

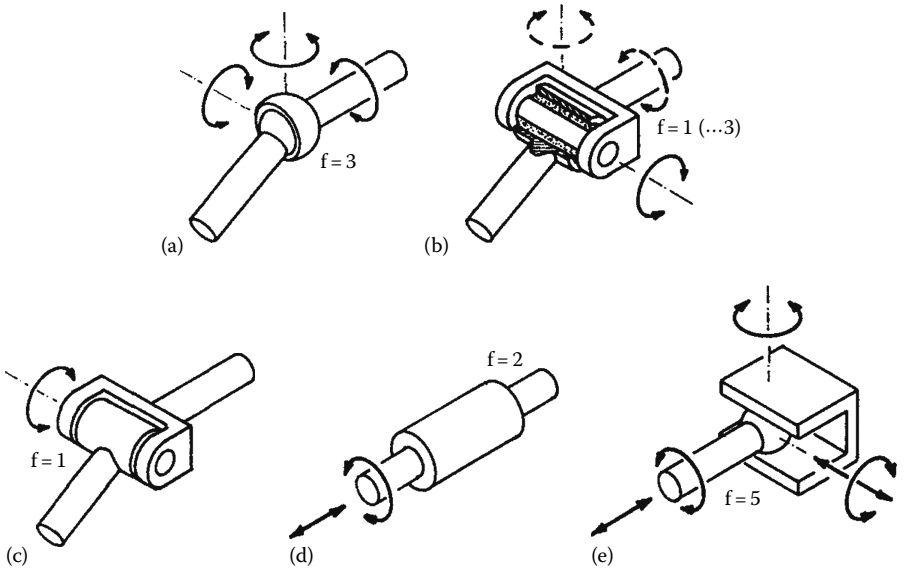


FIGURE 21.3 Joint types: (a) ball joint, (b) rubber joint, (c) turning joint, (d) turning-and-sliding joint, and (e) ball-and-surface joint.

In space, a freely movable body has six degrees of freedom. The degree of freedom of a complete mechanism is designated in the following by an uppercase "F" and that of a joint by a lowercase "f."

Figure 21.3 shows the most important types of joints used in suspensions: (a) is the ball joint that allows free motion about three independent axes of rotation, thus providing three degrees of freedom, f = 3. Since with many ball joints mainly only one rotational axis is utilized, the angular displacements about the two other axes being relatively small, a rubber joint (b) may be employed; the rubber joint, however, introduces restoring moments into the connected parts. The turning joint (c) provides f = 1 and the turning-and-sliding joint (d) f = 2. The ball-and-surface joint (e) is rare in wheel suspensions (f = 5).

These joints serve to connect the suspension links (Figure 21.4) to the body and the wheel carrier. The rod link (a) that has two ball joints, each with three degrees of freedom, provides six degrees of freedom for the mechanism. One of them, however, is the free rotation "r" of the link about its own axis. So the rod link provides five degrees of freedom for the suspension. The triangular link or "A arm" (b) provides f = 4, and the trapezoidal link (c) f = 2. The turning-and-sliding link (d) is represented in suspensions by a telescopic damper, consequently, in the special form shown here with the ball joint on the axis of the turning-and-sliding joint, enabling the rod to rotate freely (r = 1), and therefore provides f = 4 for the suspension.

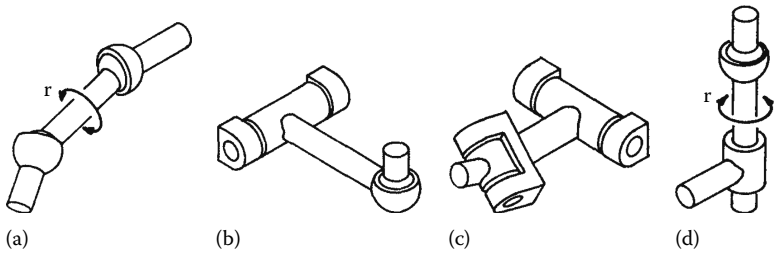


FIGURE 21.4 Types of wheel-suspension links: (a) rod link, (b) triangular link or "A arm", (c) trapezoidal link, and (d) turning-and-sliding link (r = ability of free rotation).

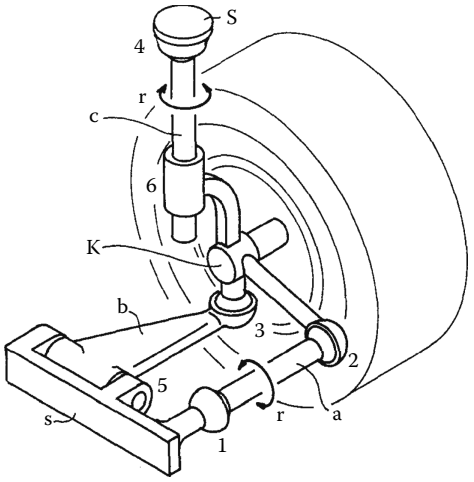


FIGURE 21.5 Schematic of an independent suspension.

Figure 21.5 shows a suspension mechanism where the most important types of joints and links are utilized. The degree of freedom of a suspension mechanism is

$$F = 6(k + l - g) - r + \sum_{1}^{g} f_i \tag{21.1}$$

where
k is the number of wheel carriers
l is the number of links
g is the number of joints
r is the number of free rotations of links
f_i is the degree of freedom of joint i

With k = 1, l = 3, g = 6, and the sum of the degrees of freedom of four ball joints, one turning joint and one turning-and-sliding joint $\Sigma f_i = 4 \times 3 + 1 + 2 = 15$ and two rotations (r = 2), the degree of freedom of the mechanism becomes F = 1, as is necessary for an independent wheel suspension.

Suspension types as used on fast vehicles are mainly classified into independent suspensions (Figure 21.6) (represented here by a simple semitrailing link system (a) and a sophisticated five-link

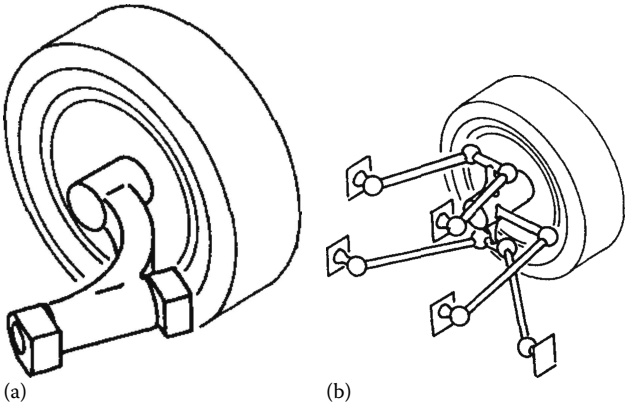


FIGURE 21.6 Examples of independent wheel suspensions: (a) Semitrailing link and (b) five-link system.

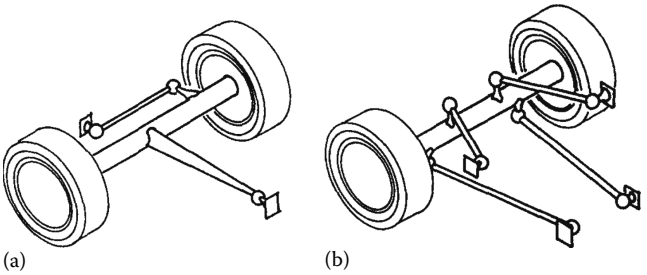


FIGURE 21.7 Examples of rigid-axle systems: (a) thrust-ball axle and (b) four-link system.

suspension (b)) or rigid-axle suspensions (Figure 21.7) (where (a) is the "thrust-ball axle" with a "Panhard rod" and (b) a four-link system). An independent suspension features one wheel and requires one degree of freedom, while a rigid axle is a wheel carrier with two wheels and, consequently, needs two degrees of freedom. Since one rod link reduces the degree of freedom of a mechanism by one, the basic model of an independent suspension is a five-link mechanism and that of a rigid-axle suspension is a system with four links.

However, there are also suspension mechanisms with two wheel carriers for two wheels but of neither the independent nor the rigid-axle type. These are the "compound suspensions" (Figure 21.8). In Figure 21.8a, the upper transverse link of each wheel of a five-link suspension is connected to the wheel carrier of the respective opposite wheel. In Figure 21.8b, the axle beam of a four-rod rigid-axle suspension has been split, and its halves have been connected by a turning-and-sliding joint; the superfluous two degrees of freedom introduced by the latter have been offset by two additional (transverse) rod links. This compound system forms the basic type of the worldwide spread family of "torsion-beam suspensions" for the rear wheels of front-drive cars (see also Figure 21.27).

For kinematic analysis of suspensions, it is popular to define "poles" and "instantaneous axes" of the mechanism and to use these for the definition of suspension properties or "suspension characteristics" like the roll center and the scrub radius or for the establishment of the balance of forces and moments at the suspension. However, this method is not generally applicable, as will be discussed in the following.

Figure 21.9 shows three variants of a simple "trapezoidal-link suspension." In Figure 21.9a, the axes of the turning joints of the trapezoidal link are parallel, and together with the upper rod link, they define an instantaneous axis of rotation m that is also parallel to the axes of the turning joints. The geometry of this suspension can be analyzed perfectly in a plane of projection that is rectangular to the joint axes. The system in Figure 21.9a represents a "planar mechanism." In Figure 21.9b, the axes of the joints at the trapezoidal link are inclined with respect to each other and intersect at a point Z. Since Z is a fixed point to the wheel carrier as well as to the vehicle body, all points of the wheel carrier rotate in space on concentric spheres about Z. The suspension in Figure 21.9b is a "spherical mechanism," and the planar mechanism of Figure 21.9a is clearly a special case of the

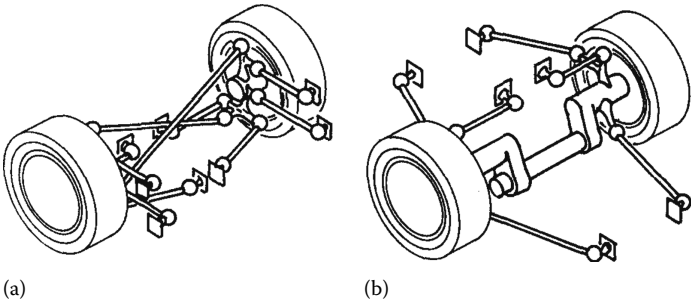


FIGURE 21.8 Examples of compound suspensions: System derivable from (a) an independant suspension and (b) a rigid axle.

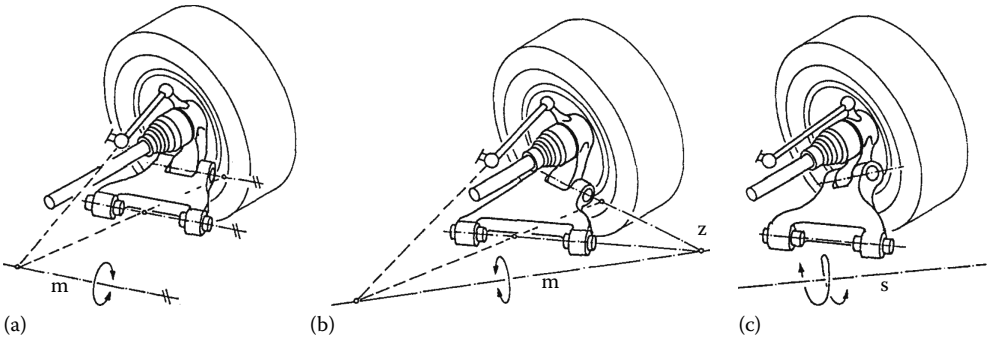


FIGURE 21.9 Three variants of trapezoidal-link suspensions: (a) Planar, (b) Spherical, and (c) Spatial Mechanism, (m = instantaneous axis, s = screw axis, z = spherical center).

spherical one, with the spherical center Z shifted to infinity. The suspension in Figure 21.9c shows skew axes of the turning joints that no longer intersect; the motion of this system can no longer be reduced to a rotation about an instantaneous axis m but is to be characterized by an instantaneous rotation about an axis s together with an overlaid axial shift along that axis. The result is an "instantaneous screw," and the mechanism is of the "spatial" type.

In suspension design, a number of "suspension characteristics" have been introduced that give information about essential properties of the suspension in connection with typical load cases or driving maneuvers. The five most important characteristics are the roll center, the braking-force support angle, the traction-force support angle, the camber change with wheel travel, and the toe-in change with wheel travel. For all these characteristics to be freely dimensioned, the suspension must be based on a mechanism defined by five parameters. An instantaneous axis of a planar or spherical mechanism is defined by four parameters (for instance, by the, respectively, two coordinates of its points of intersection with two planes), while a fifth parameter—the axial pitch—is required for an instantaneous screw.

Kinematic analysis of suspensions has, therefore, to be based on the assumption of an instantaneous screw. The axis of a screw, however, is not generally applicable for defining levers of forces or establishing the static balance of forces, in contrast to an instantaneous axis of rotation: forces that are not directed rectangularly to the screw axis generate a torque even if they intersect with that axis. Points of intersection of planes with the screw axis cannot normally serve as "poles" for kinematic or static investigations.

A very suitable and simple method for kinematic analysis of the suspension mechanism is indicated in Figure 21.10. A suspension shows five rod links and one spring element, all connected to the vehicle chassis as well as to the wheel carrier by ball joints. The (wanted) state of motion of the wheel carrier K may be characterized by the "velocity" vector $\mathbf{v}_M$ of the wheel center M as the reference point and the angular-velocity vector $\boldsymbol{\omega}_K$ of the wheel carrier. The velocity vector of any point i at the wheel carrier follows as

$$\mathbf{v}_i = \mathbf{v}_M + \boldsymbol{\omega}_K \times \mathbf{r}_i \tag{21.2}$$

For a constant length of the link $\mathbf{a}_i$, the components of the velocity $\mathbf{v}_i$ and a possible velocity $\mathbf{v}_{i'}$ at the inner joint in the direction of the link vector must be equal:

$$\mathbf{v}_i \cdot \mathbf{a}_i = \mathbf{v}_{i'} \cdot \mathbf{a}_i \tag{21.3}$$

There follows from (21.2) and (21.3)

$$(\mathbf{v}_M + \boldsymbol{\omega}_k \times \mathbf{r}_i) \times \mathbf{a}_i = \mathbf{v}_{i'} \times \mathbf{a}_i \tag{21.4}$$

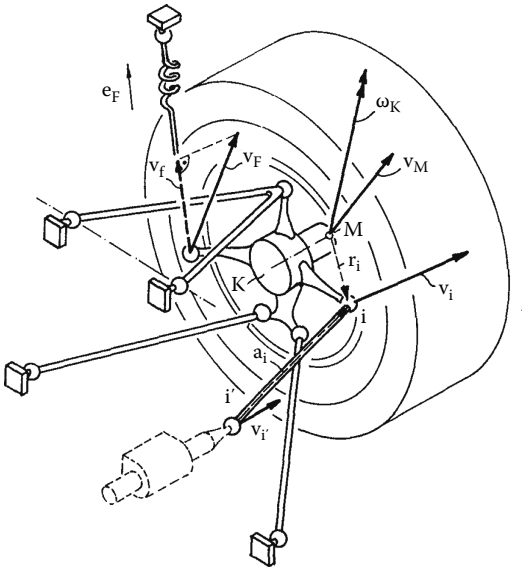


FIGURE 21.10 Motion analysis of a wheel suspension.

and—written in components—this is a linear equation. Similar equations can be established for all five rod links. The terms $\mathbf{v}_i \cdot a_i$ can represent elastic displacements of the link or also the input of a steering gear (e.g., if a_i is the track rod of a steered suspension). In all other cases, the right-hand side of Equation 21.4 is zero. The spring is a "highly compliant" rod and completes the system to a "statically determinate" structure. If compressed, its "spring compression velocity" v_f replaces the term $\mathbf{v}_i \cdot \mathbf{a}_i$ in Equation 21.4.

Six linear equations of the (21.4) type, generated by the five rod links and the spring, form a system to determine the respective three components of the vectors $\mathbf{v}_M$ and $\boldsymbol{\omega}_K$.

The "velocities" can, of course, also be regarded as "virtual displacements" if multiplied with a time differential dt.

An essential advantage of this method is that the equations serving for kinematic analysis can be also be applied to the analysis of forces and that equations for static analysis no longer need to be established. Since the "suspension characteristics" mentioned earlier are mostly derivatives of forces or displacements, the method may easily be applied to define those characteristics, as demonstrated with a simple front-wheel suspension subjected to a braking force F_{xB} (Figure 21.11).

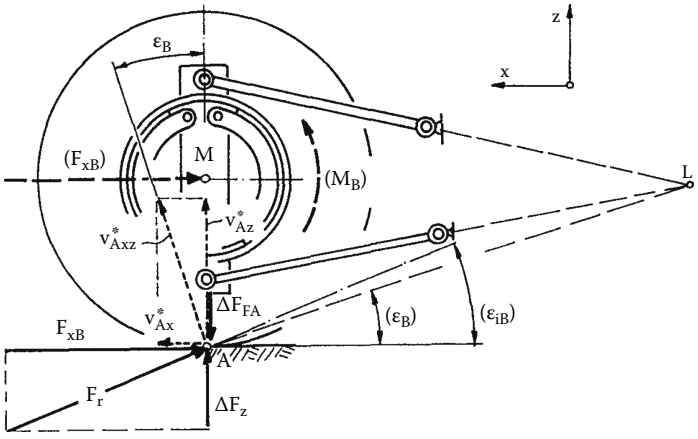


FIGURE 21.11 Balance of power (braking event).

The deceleration force causes a pitching moment, depending on the height of the vehicle's center of gravity and its wheelbase. This moment increases the front-wheel load by ΔF_Z at the expense of the rear wheel. The front-wheel brake acts between wheel and wheel carrier and represents the "torque support" that transfers the braking torque M_B to the wheel. The spring-force change (reduced to the tire contact point A) is indicated by ΔF_{FA}.

The diagram of forces is independent of the vehicle's forward velocity, so the latter can be "subtracted" from all of the "velocities" at the suspension, or, so to speak, the diagram is also valid under the assumption of a "locked" brake. Under this condition, the tire contact point A is "fixed" to the wheel carrier during braking, and, for a virtual spring compression, point A moves with a virtual "velocity" $\mathbf{v}_A^*$ rectangularly to the "polar ray" connecting it with the pole L. With the forces, and the components of the vector $\mathbf{v}_A^*$, the balance of "power" can be established: $(\Delta F_Z - \Delta F_{FA})v_{AZ}^* - F_{XB}v_{AX}^* = 0$, or with $\tan\varepsilon_B = \pm v_{AX}^*/v_{AZ}^*$ (upper sign for front wheels), we get

$$\Delta F_{FA} = \Delta F_Z - F_B \tan\varepsilon_B \tag{21.5}$$

ε_B is the "braking-force support angle" [3], a suspension characteristic, and permits the actual "anti-dive" property of the suspension to be evaluated. The spring-force change obviously diminishes as the support angle increases.

On a spatial suspension, of course, the virtual velocity vector $\mathbf{v}_A^*$ has to be determined according to Equation 21.2 assuming the torque support to be "locked."

In the rare case of inboard brakes and transmission of the braking torque via driveshafts, or in the normal case of an independent suspension with the final-drive gearbox fixed to the vehicle body, the method described earlier can be applied analogously, namely, by integrating the driveshafts into the kinematic suspension model and assuming the "torque support" (e.g., the engine) to be locked.

On a universal joint, the two shaft halves can rotate with respect to each other only about an axis that lies in the plane of symmetry Π_S (Figure 21.12). The vector of the relative angular velocity of the shaft halves therefore has to be rectangular to the "normal vector" **n** of the plane of symmetry.

For a given state of motion of a wheel carrier K of a suspension—type is immaterial—characterized by the vectors $\mathbf{v}_M$ and $\boldsymbol{\omega}_K$ in Figure 21.13, the velocity $\mathbf{v}_2$ of the outboard universal joint 2 of the driveshaft W can be determined by Equation 21.2. With $\mathbf{v}_2$, and regarding the conditions of movement at the universal joints, and assuming the inboard shaft end to be "locked," the state of motion of the driveshaft can be derived and, consequently, the relative angular velocity of the outboard

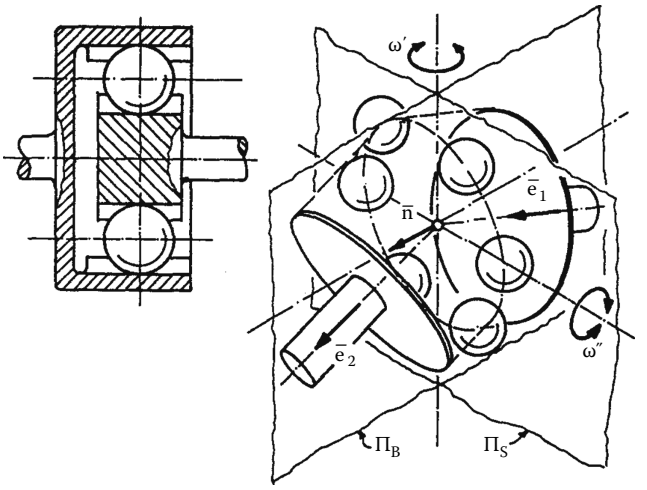


FIGURE 21.12 Homokinetic universal joint.

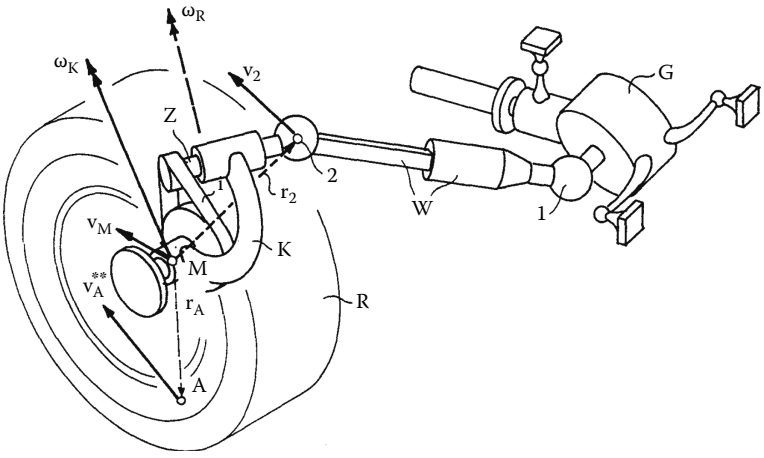


FIGURE 21.13 Wheel carrier with driveshaft and hub reduction gear.

shaft end Z of the driveshaft with respect to the wheel carrier. With this, possibly reduced by a hub reduction gear (symbolized here by a belt drive with a reduction ratio i, where i > 0 for equal direction of rotation of shaft Z and the wheel), we get the relative angular velocity of wheel and wheel carrier and so the absolute angular velocity $\boldsymbol{\omega}_R$ of the wheel. A virtual velocity vector $\mathbf{v}_A^{**}$ can be derived in analogy to Equation 21.2:

$$\mathbf{v}_A^{**} = \mathbf{v}_M + \boldsymbol{\omega}_R \times \mathbf{r}_A \tag{21.6}$$

21.3 TRACTION AND BRAKING

During acceleration or braking, the force of inertia m_{ax} acts at the center of gravity S of the vehicle (Figure 21.14) and causes the wheel-load transfer ΔF_{ZV} and ΔF_{zh} at the front (v) and rear (h) wheels. The distribution of the longitudinal traction or braking forces F_{xv} and F_{xh} at the tire contact points is determined by the design of the system, for instance, for optimum use of the friction coefficient between road and tire. The resulting forces F_r at the wheels intersect at the height of the center of gravity at a point called the "braking" or "accelerating" center E [4]. The angles ε_{iv}^* and ε_{ih}^* between the front and rear tire contact points and the center E are the "ideal" support angles; suspensions that provide these angles will achieve "100%" anti-dive or anti-squat effect upon accelerating or braking.

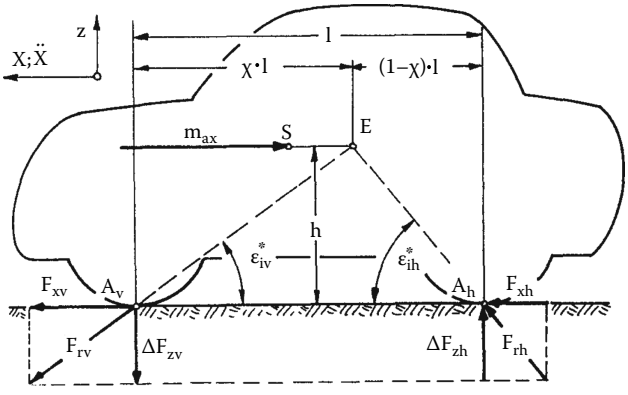


FIGURE 21.14 Diagram of forces upon accelerating.

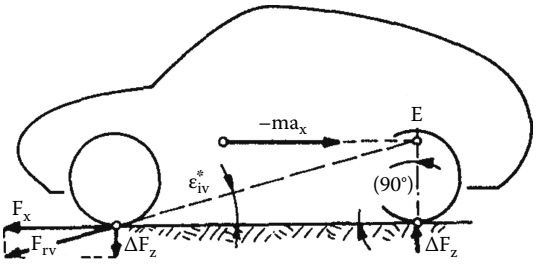


FIGURE 21.15 Single-axle drive.

With single-axle drive or braking (Figure 21.15), the accelerating or braking center E is located above the nonaffected wheel. At the latter, the "ideal" support angle is 90° that means that anti-dive or anti-squat measures are not possible. The ideal support angle of the driven or braked suspension then follows simply from the center-of-gravity height and the wheelbase of the vehicle.

As already demonstrated with Figure 21.11, the support angles follow from virtual velocity vectors at the tire contact points with virtual wheel travel, assuming the torque support (brake or engine) to be locked, and can be determined in the case of a torque support at the wheel carrier (wheel brake, or also hub engine) by

$$\tan \varepsilon^* = \frac{\pm v^*_{Ax}}{v^*_{Az}} \tag{21.7}$$

and in the case of torque transmission via driveshafts and possible hub reduction gears by

$$\tan \varepsilon^{**} = \frac{\pm v^{**}_{Ax}}{v^{**}_{Az}} \tag{21.8}$$

(upper sign valid for front wheels).

Typical cases of a torque support mounted at the wheel carrier are shown in Figure 21.16: a wheel-carrier drive unit (a) as used on light motorcycles, a "live" rigid axle with the final-drive gearbox integrated into the axle beam (b) and a brake mounted to the wheel carrier of an independent suspension (c). The torque support at the vehicle body, combined with transverse driveshafts, is typical for an independent suspension (Figure 21.17a) and for a de Dion rigid axle (b).

On steerable suspensions, it is difficult to achieve large braking-force support angles, since this is mostly coupled with a "pole" close to the wheel (like the pole in Figure 21.11), and may cause

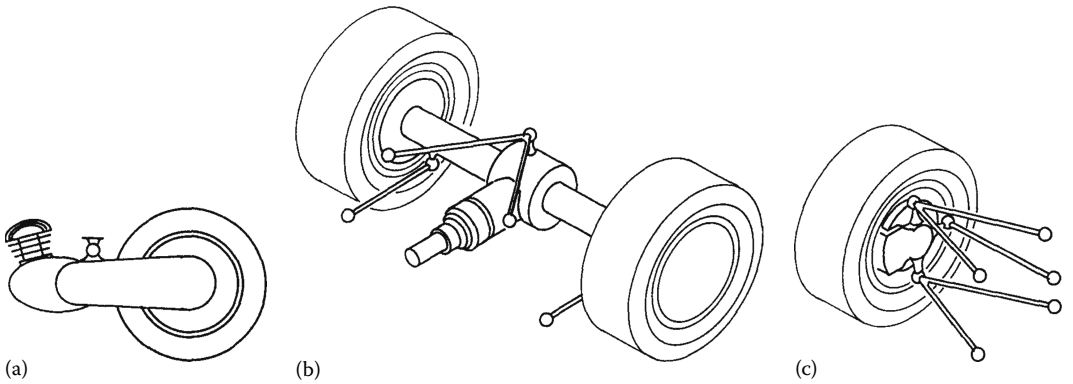


FIGURE 21.16 Torque support mounted to the wheel carrier: (a) wheel-carrier drive unit, (b) rigid axle with final drive ("live axle"), (c) brake calliper mounted to the wheel carrier of an independent suspension.

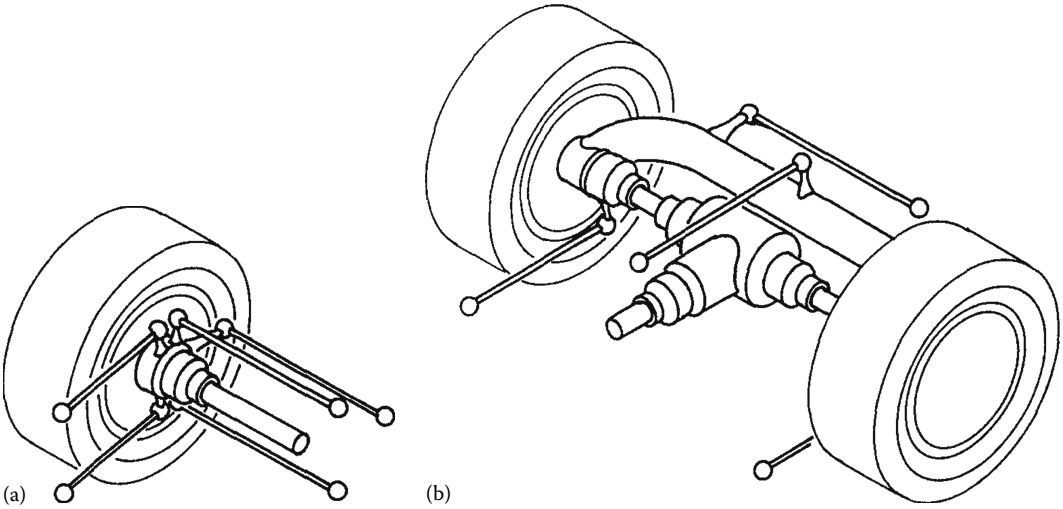


FIGURE 21.17 Torque support mounted to the vehicle body: (a) independant suspension with drive shaft and (b) De-Dion rigid axle.

undesirable change in the "castor" angle and offset with wheel travel (see Section 21.5). However, a steerable suspension with the "kingpin" arranged between the suspension and the chassis, as usual on motorcycle forks (Figure 21.18), may enable a large support angle ε^* with minimized castor change, by arranging a "leading link" K approximately at a right angle to the kingpin and controlling a rotatably mounted brake support BA by a torque reaction rod BS. The same applies to the (historical) Dubonnet suspension for passenger cars.

On an independent suspension with driveshafts, the tire contact point A will approximately copy the path of the wheel center M with wheel travel, even if the inner shaft end is assumed to be locked (Figure 21.19). Therefore, the support angle in the case of traction is comparatively low and approximately equal to the "wheel-travel angle" (see Section 21.6).

A hub reduction gear, however, may have a considerable influence on the support angle if the wheel carrier shows a rotation in side view with wheel travel. The suspension in Figure 21.20 has a

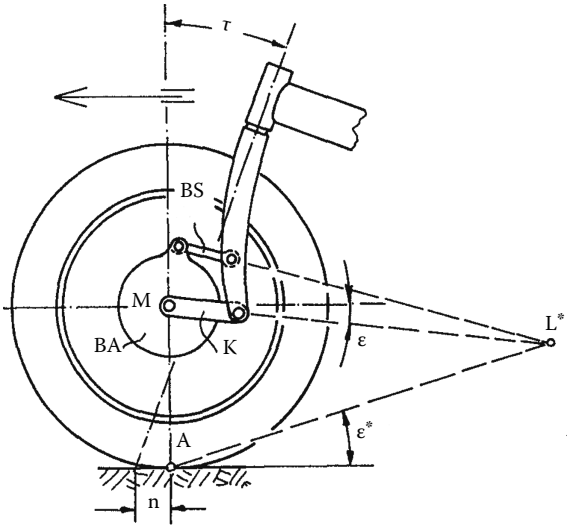


FIGURE 21.18 Rotatably mounted brake support.

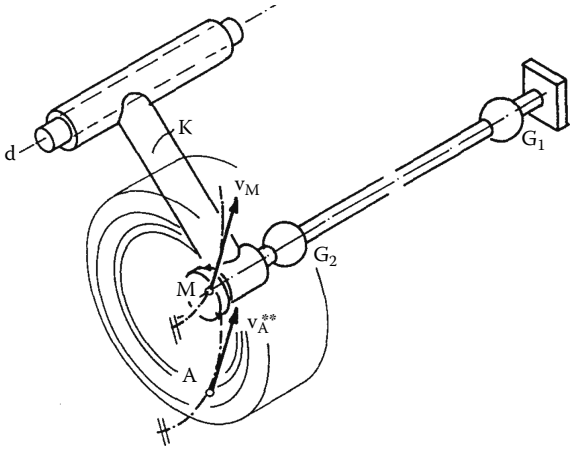


FIGURE 21.19 Independent suspension with driveshaft.

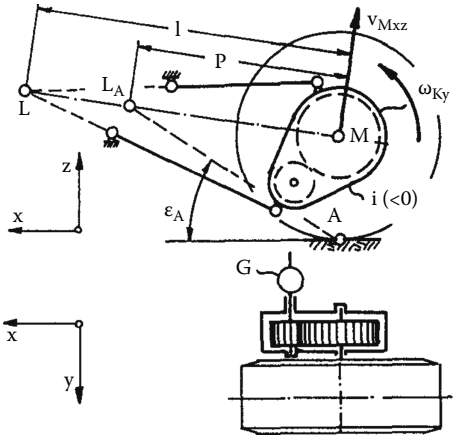


FIGURE 21.20 Suspension with hub reduction gear.

"longitudinal pole" L close to the wheel and is driven via a transverse driveshaft and a hub reduction gear. The distance of the "traction pole" L_A can be approximated by

$$p = \frac{li}{(i-1)} \tag{21.9}$$

and in the case of reversed rotation ($i < 0$), L_A is located between the suspension pole L and the wheel center M.

Overcompensation of traction or braking pitching will normally cause discomfort and will at any rate lead to increased tire wear, caused by longitudinal slip on bad road surface.

A special case of driven-independent suspension, without any universal joint in the driveline, is the rare swing-axle system with a "swinging" bevel gear (Figure 21.21). Assuming the engine to be locked, the pinion of the final drive will be locked, too. With virtual wheel travel, the swing arm will rotate about the axis d, while the wheel (as a spatial body) is fixed firstly at the point T of intersection of its axis with the axis d and secondly at the contact point E of the gears. So the virtual axis of wheel movement is the line m^{**} running through T and E and generates a pole L^{**} in the plane of the wheel and thus a considerable support angle ε^{**} (however, with suspensions of this type, the pinion at the other wheel will usually be *behind* it and create a support angle of inverted sign).

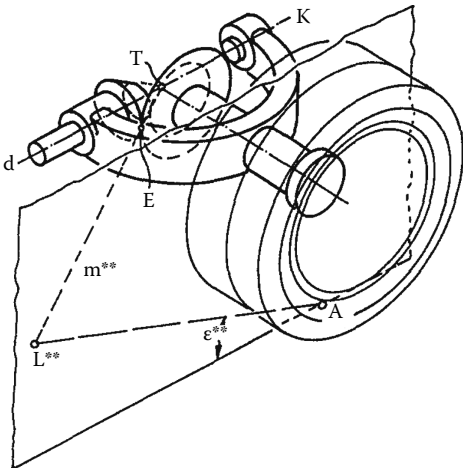


FIGURE 21.21 Swing axle with swinging bevel gear.

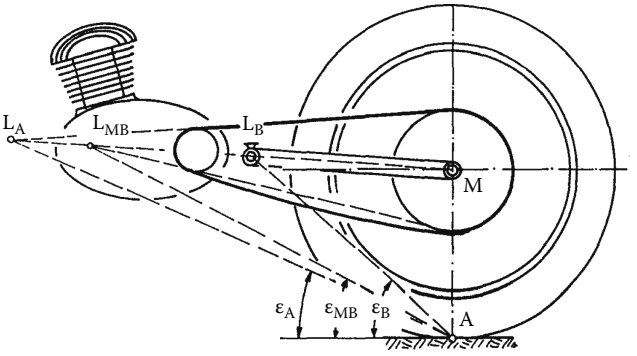


FIGURE 21.22 Support angles on a chain drive suspension.

A chain drive rear suspension with a longitudinal swing arm, as used for motorcycles (Figure 21.22), features three support angles: one angle ε_B for the brake fixed to the swing arm (the wheel carrier) and support angles ε_A and ε_{MB} for the cases of traction and engine braking that follow from the points of intersection of the swing arm and the respective chain run under tension [5].

For tandem-axle aggregates as usual on trucks and trailers, careful harmonization of the support angles is needed to avoid excessive axle-load difference upon braking or accelerating and, with that, premature slip of the axle with reduced load [1,6].

On vehicles with variable braking or traction-force distribution (for instance, motorcycles with independently operated brakes or also vehicles with the braking or traction antilock system in action), the support angles should better not exceed the "ideal" angles for the "worst case"—that is, single-axle traction or braking.

A braking power control device influenced by spring loading (monitoring the deflection of a mechanical spring or the pressure of a gas spring) cannot perceive a dynamic load change if the suspension is designed for 100% anti-pitch geometry. In such a case, the use of sensors for longitudinal accelerations will be preferable.

21.4 CORNERING

In steady-state cornering, the centrifugal force at the vehicle's center of gravity generates a roll moment that increases the outer wheel loading and decreases the inner. Figure 21.23 shows a simplified model of a vehicle, the suspensions being represented by rigid axles with "slot guidances"

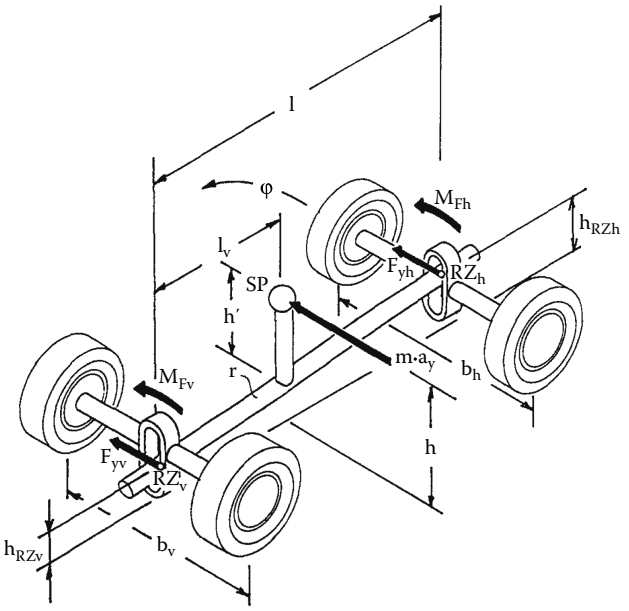


FIGURE 21.23 Forces and moments in cornering.

that allow the vehicle body to move up and down and to roll. The springs are not depicted but are represented by their restoring moments M_{Fv} and M_{Fh}. The lateral forces F_{yv} and F_{yh} are transmitted to the vehicle body via the contact points RZ_v and RZ_h. Since the center of gravity SP is located above the line r connecting the contact points, by the distance h′, the centrifugal force causes a roll moment that loads the system of springs and generates a roll angle φ. The line r is therefore called the "roll axis," and the contact points RZ are the "roll centers."

On independent and on compound suspensions, the "unsprung masses" also generate a roll moment that has to be taken up by the system of springs, depending on the gradient of camber change with wheel travel in relation to the vehicle body.

The roll moment of the vehicle body that acts on the springs is distributed to the front and rear axle according to their "roll spring rates" and causes the moments M_{Fv} and M_{Fh}. With respect to pitch oscillations of the vehicle excited by single obstacles (e.g., a railway crossing), the "natural frequency" of the front suspension is usually chosen lower than that of the rear one (see Section 21.6). Consequently, the contribution of the front springs to the "roll rate" is normally lower than that of the rear springs—which is undesirable because it results in a reduced wheel-load difference at the front wheels at the expense of the rear wheels. To compensate for this, "stabilizer" or "anti-roll" bars are used on front suspensions to reinforce the front roll rate (Figure 21.24a), and possible "transverse compound springs" (b) may additionally lower the rear roll rate in order to secure an understeering behavior of the vehicle. Within certain limits, the roll-center heights above the road surface can also serve for influencing the roll moment distribution.

Front-wheel suspensions with a significant braking-force support angle may have a further effect on wheel-load transfer because the longitudinal component of the lateral force increases with the steering angle and promotes understeering [7].

Determination of the roll center is easy at a rigid-axle suspension. In Figure 21.25, the fulcrum of the slot guidance is clearly the point of lateral-force transmission between vehicle body and axle. Determination of the roll center by application of "virtual power" (as already used in Section 21.3) has, of course, to take into consideration that the rigid axle has two degrees of freedom, one of them catering for the bounce motion and certainly not being suited for body roll analysis, firstly because it does not show horizontal velocity components at the contact points and so is unable to generate

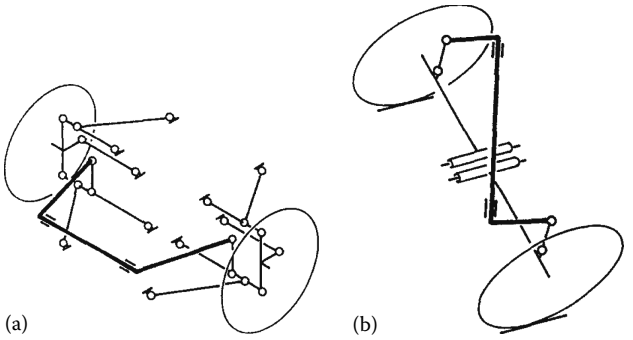


FIGURE 21.24 Stabilizer bar (a) and transverse compound spring (b).

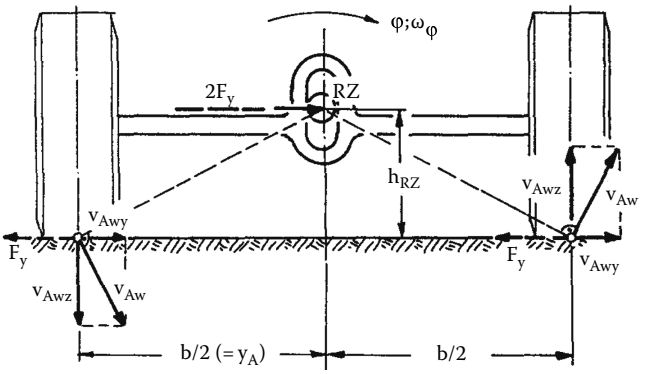


FIGURE 21.25 Roll center of a rigid axle.

"virtual power" with lateral forces and secondly—and decisive—because it is part of a "symmetrical" state of motion, while vehicle roll is an "antimetric" event. With antimetric motion of the axle, however, we get antimetric velocity vectors $\mathbf{v_{Aw}}$ at the tire contact points rectangular to the connecting lines of the contact points A and the fulcrum. The "power" of the lateral forces F_y and the velocity components v_{Awy} must be equal to the "power" of the rolling moment $2F_y h_{RZ}$ and the angular roll velocity ω_φ: $2F_y h_{RZ}\omega_\varphi = 2F_y v_{Awy}$, and since $v_{Awz} = \omega_\varphi y_A$, the roll-center height follows as

$$h_{RZ} = y_A\left(\frac{v_{Awy}}{v_{Awz}}\right) \tag{21.10}$$

With an independent suspension, there is only one form of wheel motion in relation to the vehicle body (Figure 21.26), while a compound suspension like the well-known "torsion-beam" or "H-frame" suspension in Figure 21.27 normally, of course, shows different forms of symmetrical and antimetric motion. With symmetrical up-and-down travel of both wheels, the whole "frame" rotates about the connecting line of the pivot points L and L′, that is, the axis m_p of parallel motion. With antimetric motion, however, the transverse beam is twisted about the connecting line of the shear centers T of its "open profile section," the medial point M of this line remaining "fixed" for reasons of antimetry. So the axis of antimetric motion m_w of each wheel runs through M and the respective pivot point L (or L′) and thus defines a respective "pole" Q_w in the cross-sectional plane. The antimetric velocity vector (not shown) at the contact point A in the cross-sectional plane is rectangular to the connecting line between A and Q_w and defines the roll center RZ.

The preceding considerations are valid only for very small lateral accelerations, where the lateral forces are approximately equal at both wheels and where the roll angle of the vehicle is negligible.

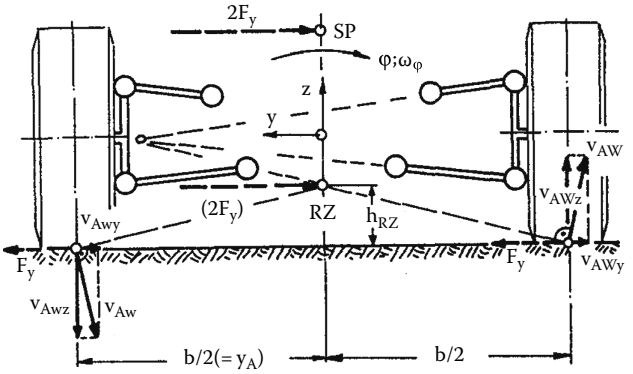


FIGURE 21.26 Roll center of an independent suspension.

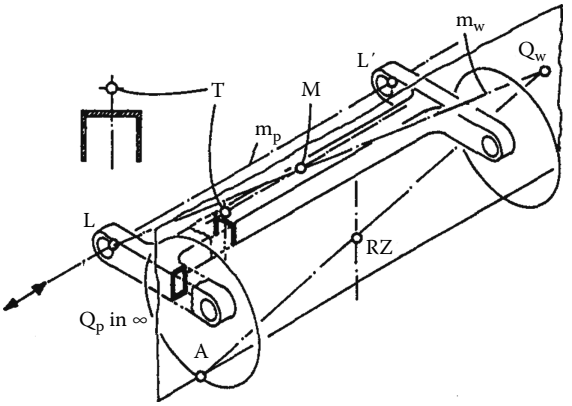


FIGURE 21.27 Roll center of a compound suspension.

With increasing lateral acceleration and roll angle, the positions of the suspensions of both wheels of an "axle" become nonsymmetrical (Figure 21.28). Here, the simple definition of a "roll center" hits up against its limits. Above all, it cannot be expected that the problem of high lateral acceleration would be solved by simple geometrical considerations, since the diagram of forces in Figure 21.28 is statically "overconstrained": the lateral forces F_{qa} and F_{qi} act in-line and result from a combination of nonlinear tire properties like the slip angle diagrams, camber forces, and toe-in or toe-out forces.

The suspension in Figure 21.28 is designed for a straightened path of the contact point A—which means "constant roll-center height above the road" and can be achieved by transverse links of unequal length, the length being approximately inversely proportional to the distance from the road [8]; see the upper sketch of Figure 21.28. The suspension is planar, so the poles can be used for the analysis of forces. Obviously, the "polar ray" A_a–P_{qa} of the outer wheel ascends considerably from the road, while the polar ray of the inner wheel is by far closer to the horizontal line. So the outer suspension can react to the vector sum of the axle-load transfer ΔF_{Ra} and the lateral force F_{qa} by a "kinematic" force K_a that reduces the necessary spring-force change ΔF_{FRa} to almost half the value of the required spring-force change at the inner wheel. The travel of the outer wheel toward bump is remarkably lower than the inner travel toward rebound; the vehicle is "jacked up."

Obviously, the jacking-up effect can be reduced by adapting the inclinations of the polar rays of both wheels to the respective forces acting so that the spring-force changes become equal on both wheels. This, of course, requires a suspension with a curved path of the tire contact

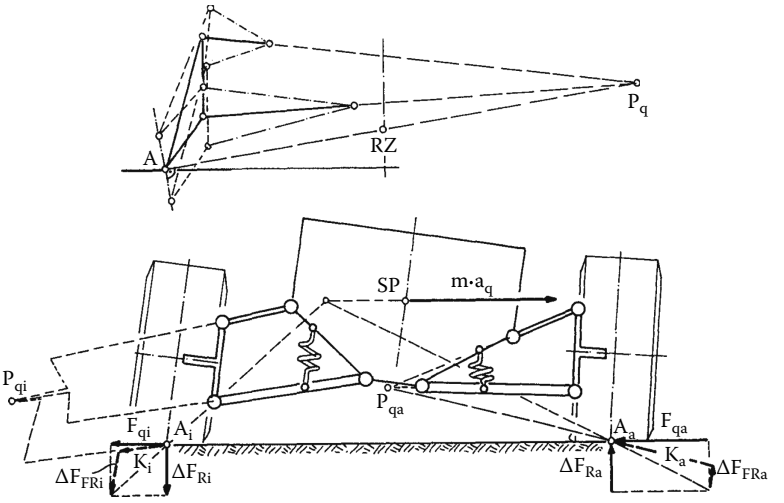


FIGURE 21.28 Independent suspension with high lateral acceleration.

point—consequently with a roll center sinking down to the road with bump. Modern independent suspensions preferably show a geometry where the roll center sinks downward with bump quicker than the vehicle body.

On rigid axles, jacking up is avoided for a roll center that appears "fixed" to the vehicle body, that is, that sinks toward the road with bump as quickly as the vehicle (this is true, for instance, for the axle shown in Figure 21.25).

Driving stability of the vehicle and its behavior in cornering can also be influenced by kinematic self-steering ("bump steer") of the suspension. The rise of a steering angle from vehicle roll can be easily explained with a rigid-axle suspension (Figure 21.29). The four-link mechanism can rotate in relation to the body only about the line connecting the points of intersection P_v and P_h of the links. Consequently, the line P_vP_h is the instantaneous axis m_w of antimetric motion and, of course, defines the roll center RZ. The "roll axis" r of the vehicle runs through RZ (see Figure 21.23). With vehicle roll angle φ and an angle κ between m_w and r, the rigid-axle beam will in plan view assume a steering angle

$$\delta = \kappa\varphi \tag{21.11}$$

similar to the axles of the well-known steerable roller skate (which employs distinctly skew axes m_w).

A positive steering angle δ on a rear axle increases the cornering radius in a left-hand corner and, consequently, promotes understeering; the contrary applies to a front axle.

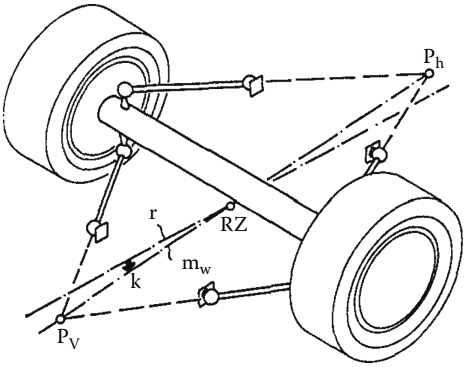


FIGURE 21.29 Four-link rigid-axle suspension.

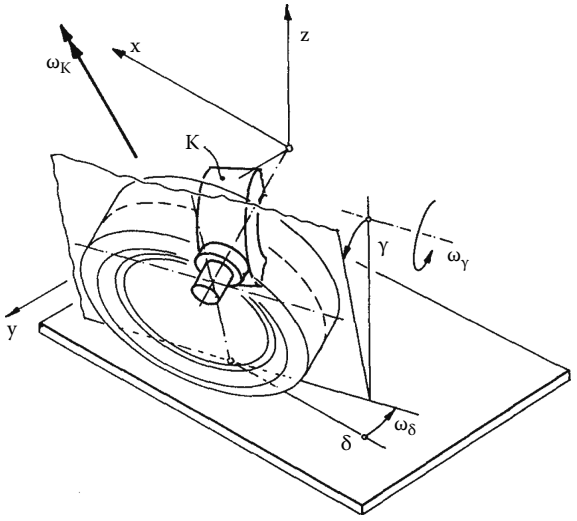


FIGURE 21.30 Angular interdependences at a wheel carrier.

With independent or compound suspensions, the definition of an angle κ is difficult and of little use because the two wheels may assume different steering angles that have to be divided into resulting toe-in angles and steering angles.

In general, a wheel carrier K with given steering angle δ, wheel camber angle γ, and angular velocity $\boldsymbol{\omega}_K$, as shown in Figure 21.30, generates a wheel camber velocity [1,6]

$$\omega_\gamma = -\omega_{Kx}\cos\delta - \omega_{Ky}\sin\delta \tag{21.12}$$

and a steering velocity

$$\omega_\delta = -\omega_{Kx}\tan\gamma\sin\delta + \omega_{Ky}\tan\gamma\cos\delta + \omega_{Kz} \tag{21.13}$$

(note that the steering velocity is not necessarily equal to the vertical component of ω_K!).

With a vertical component v_{Az} resulting from a virtual spring deflection (or, because of the same value, velocities v^*_{Az} or v^{**}_{Az}), the camber change and the steering angle with wheel travel become

$$\frac{d\gamma}{ds} = \frac{\omega\gamma}{v_{Az}} \tag{21.14}$$

$$\frac{d\delta}{ds} = \frac{\omega\delta}{v_{Az}} \tag{21.15}$$

Typical steering effects of an independent suspension are discussed using Figure 21.31. The double-wishbone suspension with a rod link (e.g., the track rod of a steerable system) behind the wheel may be designed so that the inboard joint of the rod is in the center of curvature of the path of the rod's outboard joint (position 0). The toe-in curve $\delta_v(s)$ has, therefore, a vertical tangent in the normal position (s = 0)—which means that dδ/ds is zero. With bump, however, toe-in increases slightly while, with rebound, the wheel moves progressively toward toe-out.

This "s shape" of the curve is typical for multi-link suspensions and is caused by the deviation of the theoretical path of the outboard joint of the rod, as given by the wheel-carrier motion (here with a variable radius of curvature), from the real path that is a circle about the inner joint of the rod.

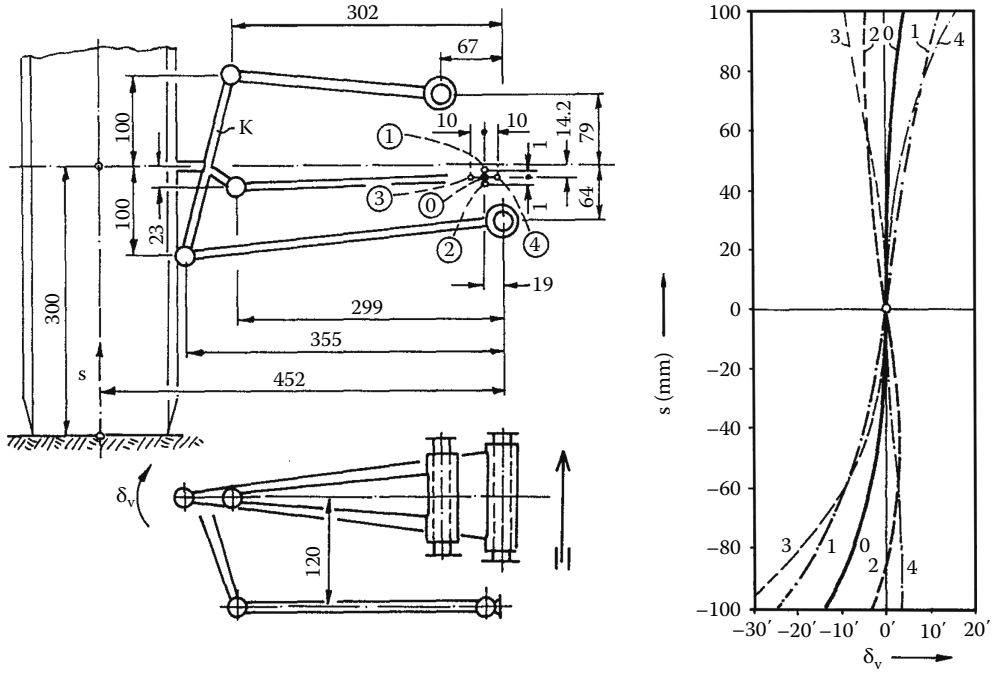


FIGURE 21.31 Toe-in against wheel travel for an independent suspension.

Displacement of the inner joint of the rod by merely 1 mm up or down (positions 1 and 2) leads to a considerable change in the toe-in gradient, the curve, however, retaining its s shape. On the other hand, lengthening or shortening the rod (positions 3 and 4) by ±10 mm without changing the height does not affect the gradient but deforms the curve by an overlaid circular shape.

These tendencies will be inverted with the rod link before instead of behind the wheel.

The diagram shows up two important things: Firstly, careful control of especially the vertical positions of joints (in production and by suitable design) is wanted to avoid undesirable tolerances of steering angles; secondly, deliberate deviation from the "kinematically correct" layout may serve for desirable steering effects or for compensation of elastic steering angles (Section 21.7).

21.5 STEERING GEOMETRY

Front wheels of fast vehicles are steered by axle-pivot (or "kingpin") steering that makes for very small lever arms of disturbing forces, saves space in the vehicle, and causes only negligible loss of embraced ground space (so secures stability against upset).

As for suspension geometry, special "steering characteristics" have also been established for steering geometry. The conventional definitions of most of these characteristics are shown in Figure 21.32.

The kingpin axis d is normally defined by two ball joints at the wheel carrier and is therefore "fixed" to the latter. It is inclined in cross section of the vehicle by the "kingpin inclination" σ and in side view by the "castor angle" τ. The horizontal distance between the intersection point D of the kingpin with the road surface and the center plane of the wheel is the "scrub radius" or "kingpin offset" r_S, while the corresponding distance at wheel-center height is the "wheel-center offset" r_C. In side view to the wheel, the horizontal distance between point D and the tire contact point A is the "castor offset" n, and the corresponding characteristic at the height of the wheel center is the "castor offset at wheel center" n_τ.

While these characteristics have already been standardized worldwide, this is not true for another, less visualized characteristic that is also of very great importance with respect to the relevant amount of the wheel load: the "wheel-load lever arm" p [1,6,7].

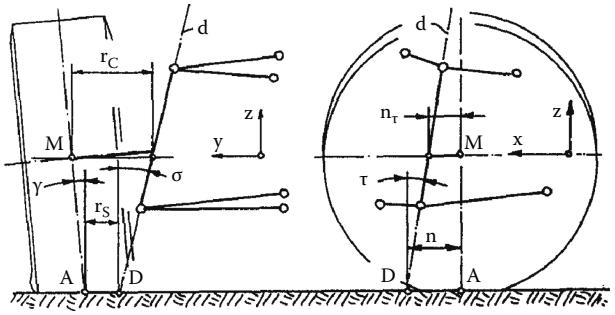


FIGURE 21.32 Conventional definitions of steering characteristics.

On a suspension with the kingpin fixed to the wheel carrier, as in Figure 21.32, the wheel-load lever arm can be calculated by

$$p = r_S \tan\tau + n \tan\sigma \tag{21.16}$$

At the end of the 1970s, steerable suspensions appeared with a "kingpin" defined by an instantaneous axis or even an instantaneous screw, to achieve more freedom in kinematic layout especially with respect to antilock braking systems and the space available in the wheels for brake disks. Figure 21.33 shows how a suspension with conventional "fixed" kingpin axis (d) can be transformed into one with a "virtual" kingpin axis (i) by splitting the triangular links into separate rod links. An early example of such a suspension is depicted in Figure 21.34. The triangular link of the strut-type design is split into a transverse and a tension link, both coupled with the wheel carrier by separate ball joints.

With such a design, the kingpin is normally variable with respect to the suspension links as well as to the wheel carrier, and several of the conventional definitions discussed with Figure 21.32 are no longer valid. Moreover, the virtual kingpin axis may become an instantaneous screw and, consequently, no longer be applicable for the definition of lever arms for forces.

For this reason, there is a need for improved definitions of the steering-geometry characteristics but, of course, in full compatibility with those hitherto. The method will be the same as used in the preceding considerations, namely, application of "virtual power balance" (Figure 21.35).

Since angular interdependences are not affected by the question of whether the kingpin axis is a screw axis or not, the definitions of the kingpin inclination σ and the castor angle τ can be retained:

$$\sigma = -\text{atn}\left(\frac{\omega_{Ky}}{\omega_{Kz}}\right) \tag{21.17}$$

$$\tau = -\text{atn}\left(\frac{\omega_{Kx}}{\omega_{Kz}}\right) \tag{21.18}$$

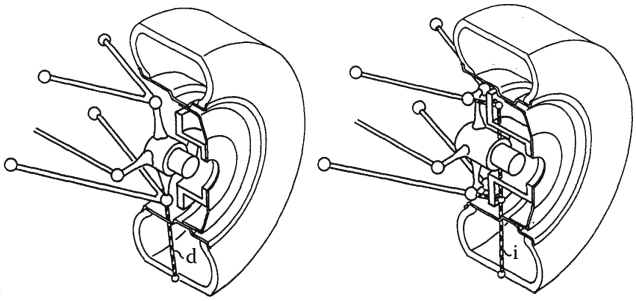


FIGURE 21.33 Suspensions with fixed (d) and virtual (i) kingpin axes.

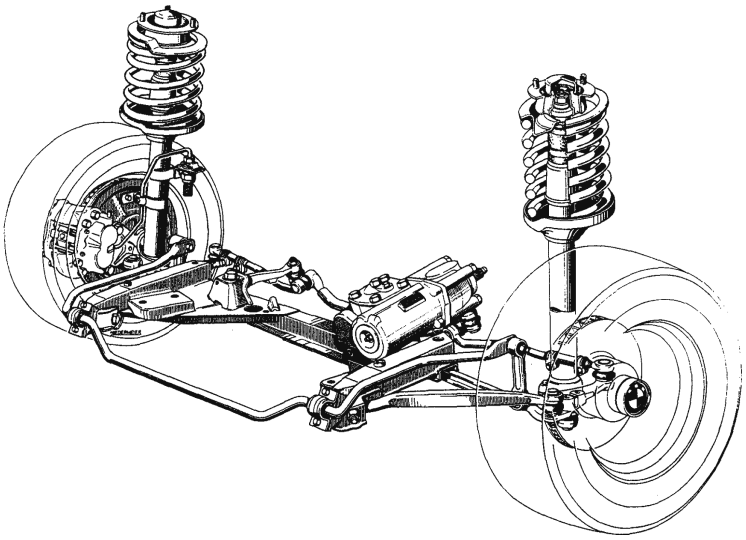


FIGURE 21.34 Strut suspension with virtual kingpin axis. (Courtesy of BMW AG, Munich, Germany, 1977.)

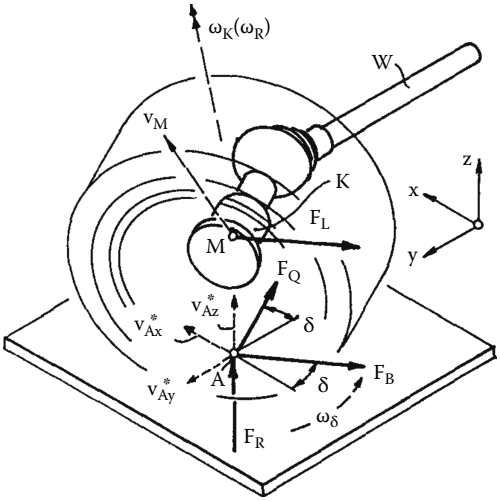


FIGURE 21.35 Velocities and forces with a steered wheel.

The scrub radius r_S is known as the lever arm of a braking force if the brake is mounted to the wheel carrier. To secure compatibility with the conventional definition, a scrub radius r_S derived from Figure 21.35, together with a braking force F_B, must give a moment $F_B r_S$ and generate a "power" $F_B r_S \omega_\delta$ with ω_δ according to Equation 21.13. With a virtual vector $\mathbf{v}^*_A$—which stems from a virtual *steering input* assuming the *spring* to be *incompressible*—the virtual "power" of a braking-force vector $\mathbf{F}_B$ is its scalar product with $\mathbf{v}^*_A$; so we get $F_B r_S \omega_\delta = \mathbf{F}_B \cdot \mathbf{v}^*_A$, and with the components $F_{Bx} = -F_B \cos\delta$, $F_{By} = -F_B \sin\delta$, and $F_{Bz} = 0$, the scrub radius is

$$r_S = \frac{-\left(v^*_{Ax} \cos\delta + v^*_{Ay} \sin\delta\right)}{\omega_\delta} \tag{21.19}$$

The wheel-center offset r_C, the castor offset n, the castor offset at wheel center n_τ, and the wheel-load lever arm p follow by similar considerations:

$$r_C = \frac{-\left(v_{Mx}\cos\delta + v_{My}\sin\delta\right)}{\omega_\delta} \tag{21.20}$$

$$n = \frac{-\left(v^*_{Ax}\sin\delta - v^*_{Ay}\cos\delta\right)}{\omega_\delta} \tag{21.21}$$

$$n_\tau = \frac{\left(v_{Mx}\sin\delta - v_{My}\cos\delta\right)}{\omega_\delta} \tag{21.22}$$

$$p = -\frac{v^*_{Az}}{\omega_\delta} \tag{21.23}$$

By multiplication with the time differential dt, Equation 21.23 changes to

$$p = -\frac{dz}{d\delta} \tag{21.24}$$

or the wheel-load lever arm is the derivative of the vehicle lift over the steering angle (with the spring assumed to be incompressible).

Typical curves of the steering characteristics γ, n, and p for suspensions with "fixed" kingpin axes and different geometrical layouts are shown in Figure 21.36.

It is clearly seen that the gradient of wheel camber γ in straight-ahead position (δ = 0) is approximately proportional to the castor angle.

The castor offset n increases with inner steering angle and decreases with outer, finally becoming negative.

The gradient of the wheel-load lever arm p for δ = 0 is the "restoring-force lever due to vehicle weight" dp/dδ, the most important prerequisite for self-centering of the steering in straight-ahead driving. On a suspension with "fixed" kingpin during a steering event with "incompressible" spring, increasing dp/dδ obviously goes along with increasing castor-offset change dn/dδ (which latter is disadvantageous for steering returnability in cornering).

This problem can be bypassed by separating the spring from the suspension links (Figure 21.37). However, the wheel-load lever arm can be influenced widely on a suspension with a "virtual kingpin."

In analogy to the generalized definitions of the steering characteristics according to Figure 21.35, but using a virtual velocity v_A^{**} for a virtual steering input with "locked" engine and shaft drive, a "traction-force radius" r_T [1,6] can be calculated as

$$r_T = \frac{-\left(v^{**}_{Ax}\cos\delta + v^{**}_{Ay}\sin\delta\right)}{\omega_\delta} \tag{21.25}$$

that is plotted against the wheel travel s in Figure 21.38 in comparison with the scrub radius r_S and the wheel-center offset r_C of a normal wheel suspension. In contrast to r_S and r_C, r_T decreases with bump and increases with rebound, thus immediately explaining the behavior of a front-drive car in

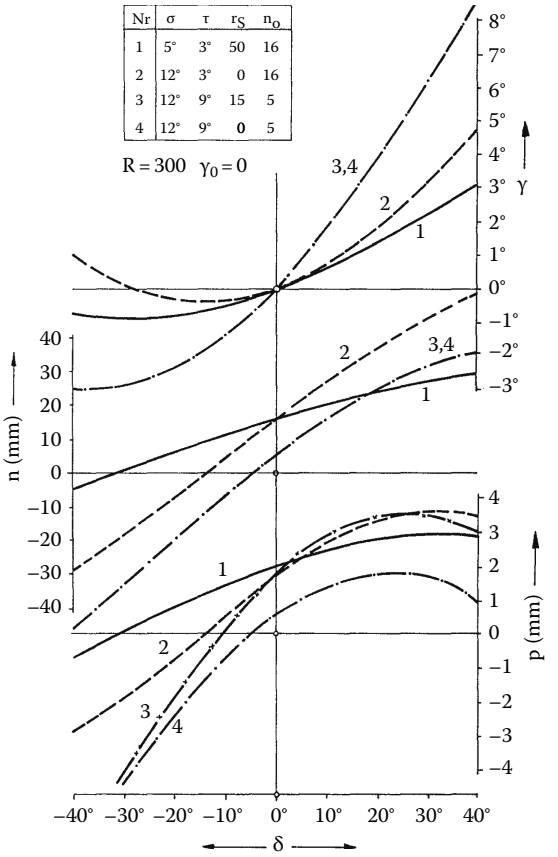


FIGURE 21.36 Camber, castor offset, and wheel-load lever arm plotted against steering angle.

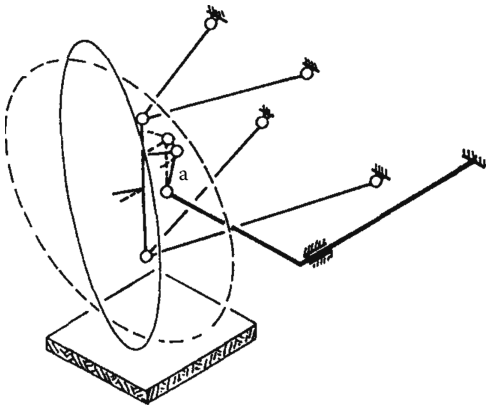


FIGURE 21.37 Free choice of restoring torque due to vehicle weight.

cornering, with the outer suspension deflected toward bump and the inner toward rebound: due to the different radii r_T, accelerating forces try to straighten the steering.

On a suspension with transverse driveshaft, but without a hub reduction gear, the radius r_T is easily determined according to Figure 21.39: the bisecting line d* of the wheel and shaft axes (or also a line parallel to it) through the intersection point H of wheel axis (a) and kingpin axis (d) intersects the road surface in a distance r_T from the tire contact point A.

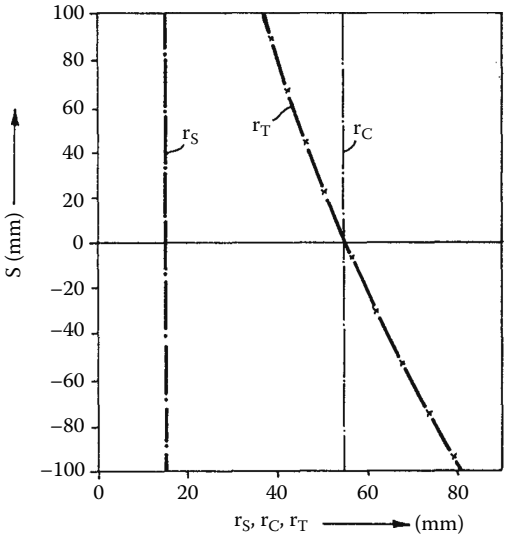


FIGURE 21.38 Traction-force radius, wheel-center offset, and scrub radius against wheel travel.

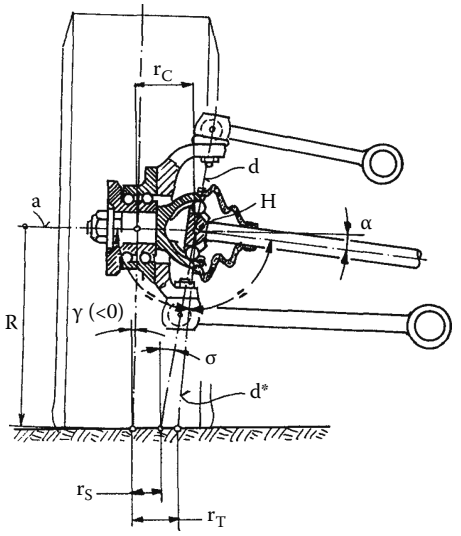


FIGURE 21.39 Graphical approximation of the traction-force radius.

With a hub reduction gear, the radius r_T can be considerably reduced, provided the kingpin inclination σ is other than zero and especially if the reduction ratio is negative (inversed rotation). With the angles in radians, r_T can be approximated [6] by

$$r_T \approx r_S + \left(\frac{R}{i}\right)\left(\frac{\sigma+\gamma-\alpha}{2}\right) \tag{21.26}$$

where

R is the wheel radius
i is the reduction ratio
σ is the kingpin inclination
γ is the wheel camber
α is the angle of the shaft deflection

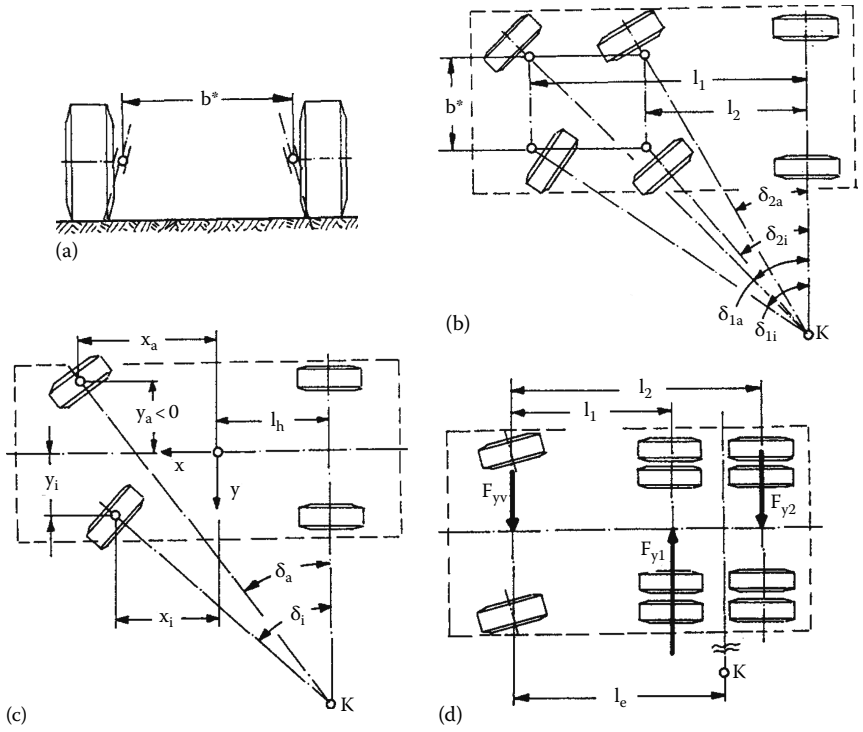


FIGURE 21.40 Ackermann condition: (a), (b) Simplifying consideration, (c) generalized and exact consideration, and (d) non-steerable tandem rear axle.

The steering linkage, consisting of levers and track rods, connects the two wheels of a steerable "axle" to the steering box and has to ensure a certain relationship of the outer and inner steering angles δ_a and δ_i. For instance, these angles may be wanted to fulfill the condition that all wheel axes of the vehicle intersect in plan view (Figure 21.40). Such a layout provides exact rolling of all wheels about a center of cornering K and is called the "Ackermann condition." The corresponding steering angles are the "Ackermann angles" (note that in vehicle dynamics, this term is applied to the medial steering angle of two wheels and is thus of quite different interpretation). The well-known simplifying "Ackermann equations"

$$\cot \delta_a = \frac{\cot \delta_i + b^*}{l} \tag{21.27}$$

for the relationship of steering angles at one axle and

$$l_1 \cot \delta_1 = l_2 \cot \delta_2 \tag{21.28}$$

for the relationship of steering angles between two axles (Figure 21.40b) are not sufficiently exact for modern suspensions with large kingpin inclinations (or even "virtual" kingpin axes) where they may lead to errors up to two angular degrees at full lock and, consequently, to serious errors in the steering-returnability layout. In such a case, it is preferable to consider the real positions of the wheels (Figure 21.40c) [1,6].

On trucks or trailers with tandem axles, the effective "wheelbase" follows from an approximation as

$$l_e = \frac{l_1^2 + l_2^2}{l_1 + l_2} \tag{21.29}$$

and is slightly longer than the medial wheelbase of the two axles (Figure 21.40d).

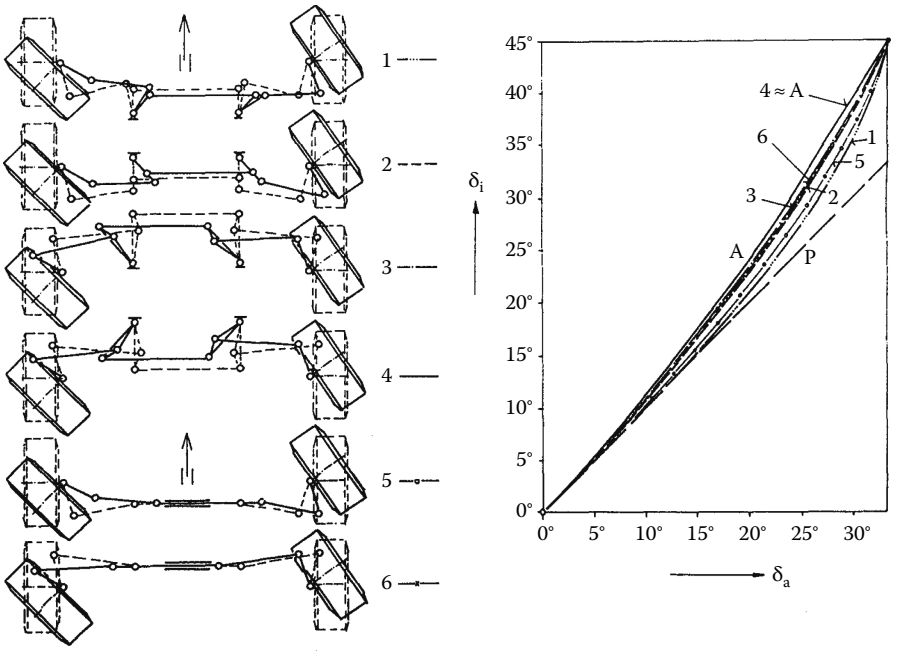


FIGURE 21.41 Typical properties of steering-linkage arrangements.

A comparison of typical properties of steering linkages for independent suspensions is shown in Figure 21.41. With respect to modern suspensions with soft longitudinal compliance (see Chapter 21.7), track rods are preferably arranged in transverse direction to avoid induced steering angles with fore-and-aft compliance.

The Ackermann curve A is not curved steadily in the same sense but tends toward a point of inflection. Only the linkage 4 is able to provide such a behavior [9] but this, however, at the expense of an extreme forward position of the steering box. The linkage 1 shows the greatest deviation of the Ackermann curve and close approximation to the curve P of parallel steering of both wheels over a wide range.

Steering returnability with high lateral accelerations is mainly secured by the lateral forces of the tires in connection with the steering-geometry characteristics.

When rolling forward, and loaded with a lateral force (Figure 21.42), a tire shows an angular deflection of its contact patch by the slip angle α, caused by the trapezoidal diagram of lateral shear τ. The resulting force of this diagram is therefore displaced behind the wheel center by the "pneumatic trail" n_R.

Consequently, the lateral force of the tire acts on a steerable suspension by an effective lever that is the sum of the (geometrical) castor offset n and the pneumatic trail n_R.

With increasing lateral force (and increasing slip angle α), however, the shear diagram (τ) begins to collapse at its rear end, and the pneumatic trail n_R decreases toward zero.

Moreover, with increasing steering angle, the castor offset n of the (dominant) outer wheel decreases and may in the end assume a negative value; see Figure 21.36.

For this reason, the restoring torque at the steering wheel usually shows a "degressive" curve with respect to steering angle and lateral acceleration. This has the advantage of avoiding high steering-wheel forces and cramped muscles of the driver's arms, securing better sensitivity for steering corrections that may be needed at the vehicle's dynamic limits. At any event, a restoring torque greater than zero is necessary to provide correct information about the actual state of forces.

The increased wheel load of the outer wheel acts on the corresponding wheel-load lever arm p and contributes to returnability if p is negative—which is true only for greater steering angles (see Figure 21.36).

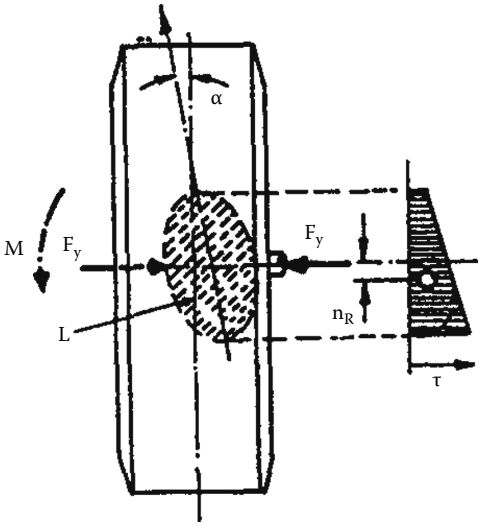


FIGURE 21.42 Tire slip angle α and pneumatic trail n_R.

A strong braking-force support angle, especially if progressive with bump, causes a "stabilizer" effect at the front axle that may lead to self-reinforcing understeer and so may impair the returnability, as already mentioned in Section 21.4.

In contrast with cornering at high lateral accelerations, the steering returnability at very low speed, with a lateral acceleration near zero—for example, during a parking maneuver—is influenced by a considerably greater number of parameters. In the extreme case, wheel-load transfer and dynamic lateral forces vanish.

If the steering geometry shows a deviation from Ackermann, for instance, because of the special property of the steering linkage (see Figure 21.36) or also because of deliberate approximation to the parallel steering curve (P) for better use of the sideslip potential of the two wheels (as very usual in the past), the angles of deviation act like considerable "toe-in" angles and generate lateral forces directed toward the medial plane of the vehicle on *both* wheels. Due to the decrease in the outer castor offset n with respect to the inner, with increasing steering angle (see Figure 21.36), these "toe-in forces" cause a moment that tends to increase the steering angle and so to impair the return of the steering. The vehicle load, of course, acts on the wheel-load lever arm in a restoring sense, but it is usually too small to counteract the "toe-in" moment.

However, with great steering angle and small cornering radius, the tire generates a "scrub torque" that tries to resist the circular path of the contact patch on the road, and this torque will in many cases "save" the returnability.

The returnability of the steering at low speed obviously depends on the difference between two large moments that, moreover, are affected by considerable tolerances caused by tire wear, pressure, temperature, manufacturer, etc. (a worn tire with abrased tread shows significantly higher sideslip forces than a well-grooved one). For this reason, good approximation of the steering geometry to Ackermann may be recommendable on suspensions with great castor-offset change, and road tests should cover the whole field of possible conditions like worn tires, high and low friction coefficients, and wrong tire pressure.

21.6 SPRINGS AND SUSPENSION

The primary purpose of springing and damping is to protect the vehicle body from impact forces resulting from road irregularities. Upon being compressed or extended, the spring stores or releases energy, the vehicle body experiencing only the spring-force fluctuation.

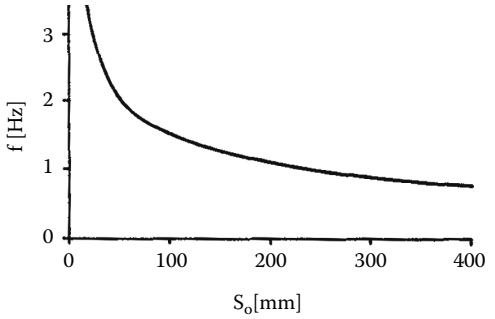


FIGURE 21.43 Natural frequency versus static deflection for a linear spring.

Most acceptable frequencies for a human being range between f = 0. 7 and 2.0 Hz that correspond to static compressions s_o of a linear spring of about 500–60 mm (Figure 21.43). Since the wheel travel of a suspension is limited to about ±100–120 mm, a soft spring cannot store much energy. A progressive spring characteristic is achieved by the superimposition of a basic and an auxiliary spring (preferably of the rubber-elastic type).

For a comfortable vehicle, the pitching frequency should not be higher than the bounce frequency. On a vehicle with long wheelbase or low pitching moment of inertia, this may require application of a "longitudinal compound springing system" in order to reduce the pitching frequency (this is typical, e.g., for light front-drive cars).

Long-wave obstacles like railway crossings excite pitching oscillations due to the delay of time between the impacts on the front and on the rear wheels. For quick fading of the pitch angle, it is usual to give the rear suspension a chance to "catch up" on the oscillation of the front suspension by providing the bounce frequencies at the front axle lower than at the rear axle. This measure, on the other hand, is contradictory to the desire for a "front roll rate" higher than the rear rate (see Section 21.4) and is the reason for the application of antiroll bars at the front suspensions.

Several features that concern the interaction of springs and suspension are taken into consideration in the following.

Even today, leaf springs are in use also still for suspension purposes. A nonsymmetrical leaf spring (Figure 21.44) shows a rotation of the axle beam connected to it about a "pole" p, with a polar distance

$$p = \frac{l_1 l_2}{(l_2 - l_1)} \tag{21.30}$$

This pole corresponds absolutely to the "kinematic" pole of a suspension mechanism and is therefore applicable in the same sense, for example, for the determination of a support angle.

Under traction or braking torque M, a leaf-sprung axle "winds up" by an angle α' (Figure 21.45); the windup spring rate is

$$c_{\alpha'} = \frac{dM}{d\alpha} = c l_1 l_2 \tag{21.31}$$

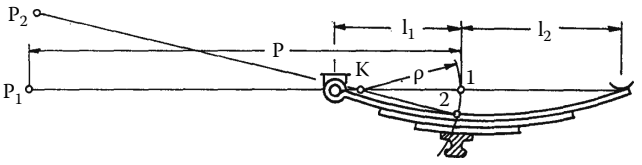


FIGURE 21.44 Nonsymmetrical leaf spring with its pole.

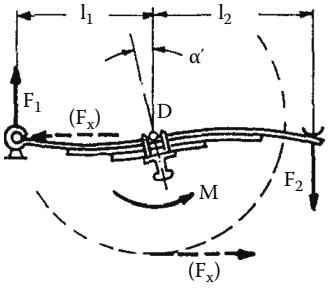


FIGURE 21.45 Windup of a leaf spring under torque.

or $c_{\alpha'} = cl^2$ for a symmetrical spring with $l_1 = l_2 = l$—which means that the resistance to windup increases proportionally to the square of the length (a decisive reason for the application of long leaf springs).

A coil spring is usually mounted between the body and a suspension link and is thus subjected to bending with wheel travel (Figure 21.46). Minimum bending moments, optimum security against buckling, and minimum height of installation can be achieved by a symmetrical deformation of the spring at both ends (Figure 21.46b).

In a suspension, the spring is usually displaced from the wheel load line, and so shows a "spring ratio" between its deflection and the wheel travel (Figure 21.47). The effective spring force F_{FA} "reduced" to the tire contact point A follows from the spring force F_F by the balance of power: $F_F v_f = F_{FA} \cdot V_{Az}$, or with the spring ratio

$$i_F = \frac{v_f}{v_{Az}} \tag{21.32}$$

we get

$$F_{FA} = F_F \cdot i_F \tag{21.33}$$

The effective "spring rate" $c_{FA} = dF_{FA}/ds = d(F_F i_F)/ds$ reduced to the tire contact point A depends on the variable spring force F_F and on the spring ratio i_F that is also variable in most cases. Therefore, the effective spring rate has to be calculated [10] by

$$c_{FA} = c_F i_F^2 + F_F \left(\frac{di_F}{ds} \right) \tag{21.34}$$

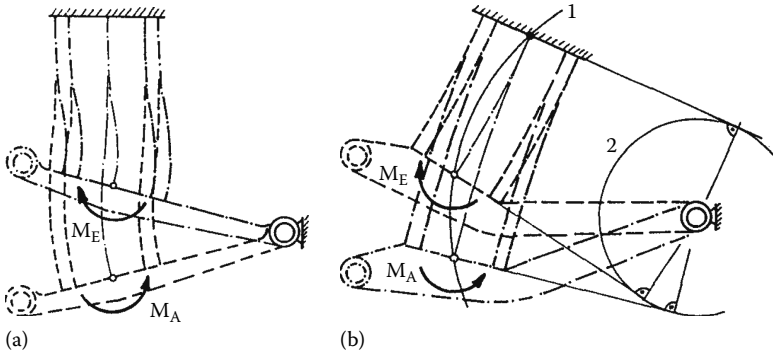


FIGURE 21.46 Coil springs and suspension links.

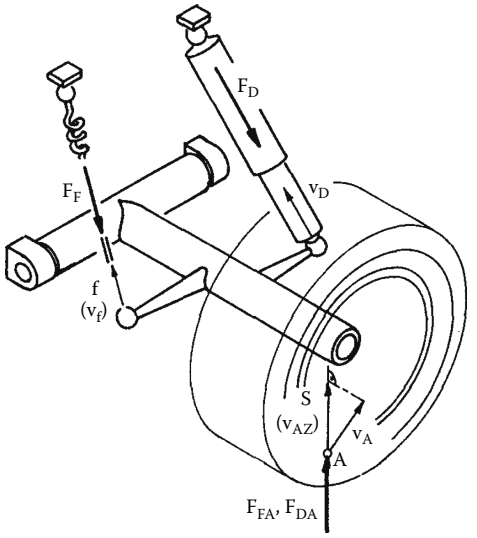


FIGURE 21.47 Spring ratio.

The second term of this equation is the "kinematic spring rate," and its application is conspicuous especially at the rear suspensions of motorcycles in order to achieve a highly progressive spring characteristic using a linear spring element.

Variable spring ratio, however, is closely coupled with variable levers and varying directions of forces. Modern wheel suspensions are designed for elasto-kinematic function—see Section 21.7—and especially for soft longitudinal compliance. To avoid fluctuating horizontal components of forces on such a suspension, kinematic tricks to influence the spring characteristic are not recommendable. The spring and the damper should preferably retain an approximately vertical position in the vehicle.

On a suspension with traction or braking anti-pitch design, and with the support angle variable with wheel travel, the longitudinal force influences the temporary spring rate during a braking or accelerating event [1,6].

Finally, another "suspension characteristic" may be mentioned, which, however, is of subordinate importance on wheel suspensions for motorcars, namely, the "wheel-travel angle" ε (Figure 21.48), by which the path of the wheel center M is inclined with respect to the vertical line in side view:

$$\varepsilon = -\text{atn}\left(\frac{v_{Mx}}{v_{Mz}}\right) \tag{21.35}$$

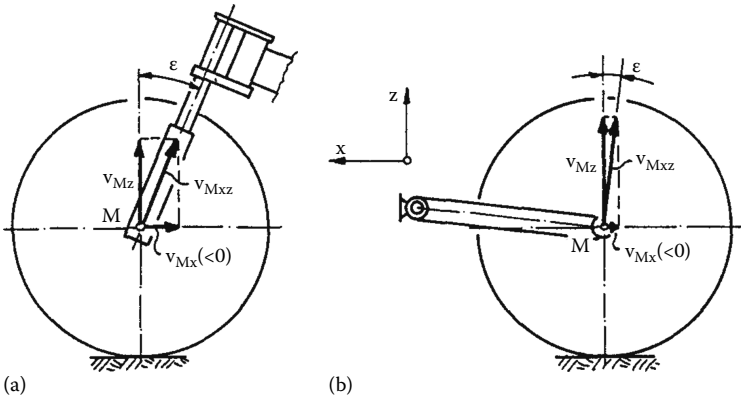


FIGURE 21.48 Wheel-travel angle at a telescopic fork (a) and a trailing link (b).

It is conspicuous for a telescopic front fork of a motorcycle (a)—and may there be useful for off-road riding—but is of minor interest on a motorcar suspension (b) having regard to the rubber-elastic mountings of the latter. It has already been mentioned in Section 21.4 that there is a relationship to the traction-force support angle of a suspension with driveshaft.

21.7 ELASTO-KINEMATICS

Road irregularities as well as tire nonuniformity excite suspension vibrations that are softened by rubber mountings. These also serve for noise insulation and have the additional advantages of being maintenance-free, friction-free, capable of recovering after temporary overload, and cost saving.

In former times, rubber mountings normally worsened the driving behavior, since they provoked undesirable elastic properties at the suspension mechanism, so it was usual to replace them by stiff (e.g., metallic) joints for motor sport purpose.

On comfortable vehicles, longitudinal compliance of a suspension may achieve as much as 15 mm. Suitable countermeasures are required to avoid unfavorable steering angles on such a suspension.

The term "elasto-kinematics" means a conscious harmonization of the spring rates of the suspension joints (and possible elastic rates of any chassis elements) and of the spatial arrangement of the suspension links to compensate for the elastic displacements that occur under external loads or even to convert them into "wanted" displacements.

A historical solution of a longitudinally compliant suspension, first applied as early as in 1933, is shown in Figure 21.49. The double-wishbone system can rotate about an axis d at the chassis against the restoring force of a rubber mounting G. The track rod Sp runs roughly parallel to the plane defined by the axis d and the kingpin axis, thus allowing longitudinal displacement without noticeable toe-in change.

This is, of course, not yet an "elasto-kinematic" assembly in the modern sense, but there is already an intention to be seen of avoiding elastic "windup" of the wheel carrier under braking force.

On a rubber joint, different spring rates in different directions are easy to achieve (Figure 21.50). A skew force F leads to a displacement f that is deflected toward the softer spring in direction "1" (the drawing also shows the—exact—graphical design according to the Mohr/Land method to determine the displacement $f = F/c^*$). These properties of rubber springs are frequently utilized especially for subframe mountings.

Elasto-kinematic investigations can be divided into two main fields: the suspension mechanism and the structure of subframes or the vehicle body. While suspensions are generally statically

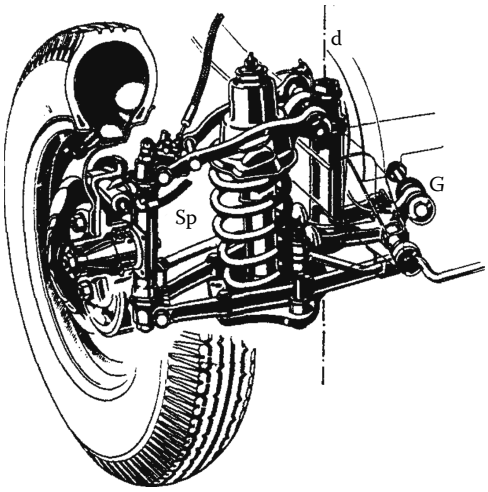


FIGURE 21.49 Early solution of longitudinal compliance. (Courtesy of Daimler Chrysler AG, Stuttgart, Germany.)

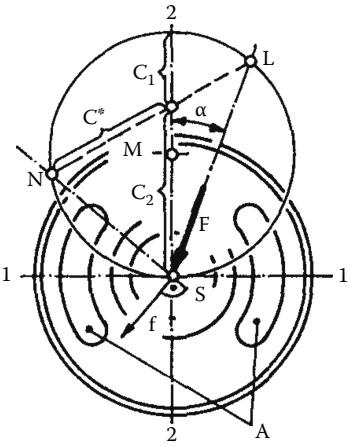


FIGURE 21.50 Rubber spring with skew loading.

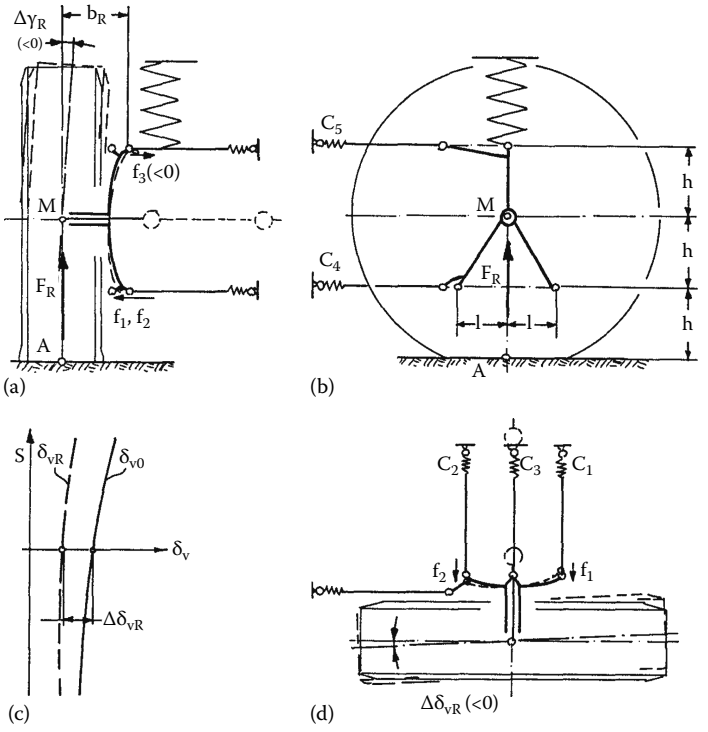


FIGURE 21.51 Independent suspension with wheel loading F_R: $c_1 \ldots c_5$ = Spring rates of the links; γ = camber angel; δ_v = toe-in angle.

determinate systems and therefore easier to survey, the deformations of spatial structures can normally be evaluated only by establishment of finite-element models.

In the first place, typical measures of elasto-kinematic harmonization of the suspension mechanism may be demonstrated using an example of a five-link system with very clear and symmetrical dimensions (on a realistic suspension, all the effects discussed here will appear in superimposition).

The wheel load F_R (Figure 21.51) is caught up by the outboard joint of the upper transverse link that supports the spring. The moment $F_R b_R$ loads the upper transverse link by compression and the two lower transverse links by tension, while the two longitudinal links are not involved.

For the vehicle's driving stability, preferably an elastic steering angle (toe-in change) $\Delta\delta_v$ is of interest and occasionally an elastic camber change $\Delta\gamma$, too. On the simple mechanism of Figure 21.51, the elastic toe-in change and camber change follow from the deformations f_1 and f_2 of the two lower transverse links and f_3 of the upper one as

$$\Delta\delta_v = \frac{(f_1 - f_2)}{2l} \quad \text{and} \quad \Delta\gamma = \frac{\left[f_3 - (f_1 + f_2)/2\right]}{2h}$$

where $f_1 = F_1/c_1$, etc. (>0 for tension force). With the elastic deformations $f_1 \ldots f_3$ we get an elastic chamber angle (here negative) $\Delta\gamma_R$ and an elastic toe-in-charge $\Delta\delta_{vR}$ caused by the wheel load force F_R.

While F_R causes a negative camber change, the toe-in angle will be negative (i.e., toe-out) for $c_1 > c_2$. Based on the "kinematic" toe-in curve δ_{v0} of the (rigid) suspension mechanism (Figure 21.51d), the "elastic" curve under wheel load δ_{vR} is displaced in normal position (s = 0) by $\Delta\delta_{vR}$ and, since the wheel load increases with bump, will show increasing offset from the curve δ_{vo}.

If undesired, the offset $\Delta\delta_{vR}$ can be compensated for by adjusting one of the lower transverse-link joints. To correct the changed gradient of the curve δ_{vR}, a deliberately "wrong" kinematic adjustment of the vertical coordinate of one of the joints of the lower transverse links is useful (see Figure 21.31), and a variation of the length of one of the links may help to correct a possibly unfavorable change in the radius of curvature of δ_{vR}.

These elasto-kinematic measures obviously allow to achieve any desired function of the suspension under wheel load. With this state of the system, the next case to be investigated may be that of lateral force (e.g., cornering) (Figure 21.52).

The lateral force F_Q loads the suspension at a point displaced behind the tire contact point A by the "pneumatic trail" n_R (see Figures 21.42, 21.52b and d) and, consequently, the lower transverse rear link by a considerably greater reaction force than the lower tansverse front link, and causes elastic compression of these two links and deformations f_1 and f_2. The upper transverse link is loaded by a tension

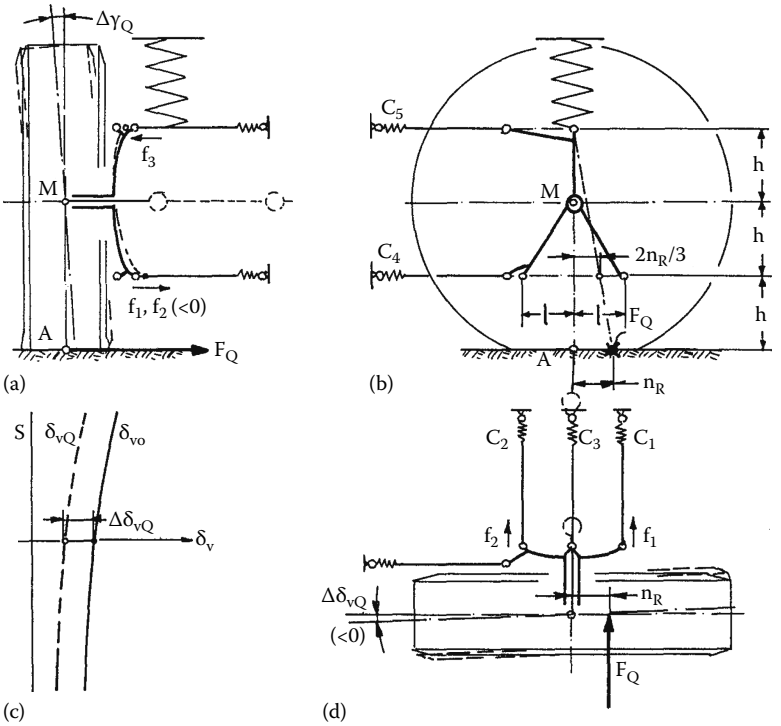


FIGURE 21.52 Suspension of Figure 21.51 loaded with a lateral force F_Q.

force and, therefore, elongated by f_3. The deformations f_1 and f_2 lead to a toe in change $\Delta\delta_{vQ}$ (positive or negative, depending on the spring rates of the lower transverse links) caused by the lateral force, and, together with the elongation f_3 of the upper transverse link, to a (positive) elastic camber change $\Delta\gamma_Q$.

On a front wheel, toe-out will be welcomed because this will improve the vehicle's behavior by an understeering reaction to a steering input (important for the driver's safe "feeling"), while on a rear wheel, toe-out means oversteering and is at any event undesirable. To compensate for undesired toe-out, there are obviously two imaginable measures: firstly, a reinforcing of the spring rate c_1 (or reduction of c_2) without affecting the suspension geometry and, secondly, a shift of the lower transverse links in fore-and-aft direction to provide more suitable reaction forces of the links. These measures, of course, require repeat of the investigations into wheel-load elasto-kinematics as discussed with Figure 21.51.

As already noted in Section 21.5, the pneumatic trail n_R decreases as the lateral acceleration increases. For this reason, any compliant front suspension will show a tendency toward elastic oversteering and any rear suspension a tendency toward elastic understeering with increasing lateral acceleration.

With the brake arranged between wheel and wheel carrier, as usual, a braking force F_B acts at the tire contact point A (Figure 21.53) and causes reaction forces at the longitudinal links, namely a tension force at the lower link and a compression force at the upper one, and elastic deformations f_4 and f_5 of them. Moreover, due to the lateral offset b_B of the longitudinal links from the central plane

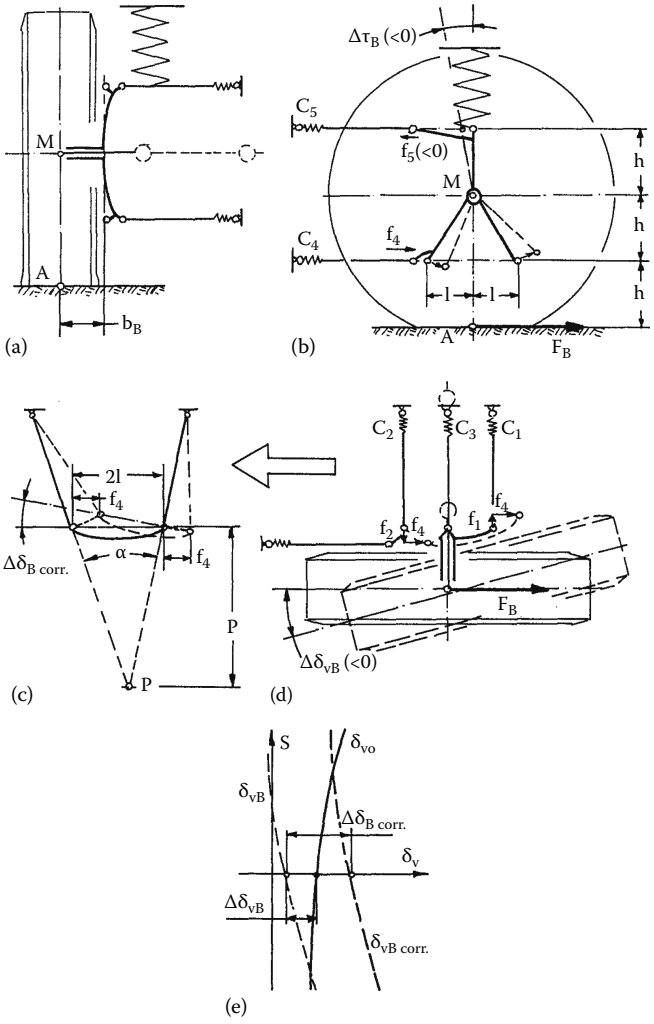


FIGURE 21.53 Suspension of Figure 21.51 loaded with a braking force F_B: τ = windup angle of wheel carrier.

of the wheel (Figure 21.53a), the lower transverse rear link is compressed by f_1, while the front one is elongated by f_2. This causes an elastic steering angle $\Delta\delta_{vB}$ that shifts the "kinematic" toe-in curve δ_{v0} in toe-out sense. While elastic camber change is here of little interest, a negative toe-in change will, of course, be favourable at a front suspension since, in cornering, the (dominant) outer wheel will promote an understeering reaction during braking. On a rear wheel, however, a toe-in increase during braking when cornering is urgently desired in order to counteract the oversteering "power-off effect" that occurs especially in emergency situations.

With the spring rates of the transverse suspension links already adapted to the preceding load cases ofwheel load and lateral force, the inversion from toe-out into toe-in requires application of a new trick (Figure 21.53c): the inclination of the two lower transverse links toward each other in plan view by an angle α to create a "kinematic" steering angle about the "pole" P under the influence of the longitudinal deflection f_4 of the lower longitudinal link (for toe-in under lateral force, this pole has of course to be displaced behind the wheel center by at least $2n_R/3$, see Figure 21.52). It may, for instance, be desired to invert the toe-out angle $\Delta\delta_{vB}$ into a toe-in angle by superimposing a correcting angle $\Delta\delta_{Bcorr}$ greater than $\Delta\delta_{vB}$. The required pole distance follows then from $p : f_4/\Delta\delta_{Bcorr}$.

The toe-in change is, however, not the only serious effect that occurs during braking. Another deformation is a considerable "windup" $\Delta\tau_B$ caused by the lengthening of the lower longitudinal and the compression of the upper link. It is obvious that the outboard joints of the two lower transverse links are displaced vertically in opposite directions, causing a remarkable change in the gradient of the toe-in curve (see the considerations with Figure 21.31). So the resulting toe-in curve δ_{vB} is determined by a toe-out displacement $\Delta\delta_{vB}$ in normal position, compensated for, of course, by a toe-in angle $\Delta\delta_{Bcorr}$, and an inclination of the curve toward toe-out with bump. The latter is tolerable at front and rear wheels, since in cornering, it promotes understeering at front as well as at rear wheels under the influence of a braking force.

Assuming the suspension to be driven by a transverse driveshaft, a fourth case of external load remains, the traction force, that normally loads the wheel carrier at the wheel center M (Figure 21.54). Due to the simplified geometry of the suspension chosen for the examples, the traction force F_T compresses the two longitudinal links by equal forces, and elastic "windup" will be negligible. With the lateral distance b_T between M and the longitudinal links in cross view, the traction force causes an elongation f_1 of the lower transverse rear link and a compression f_2 of the front one, and these deformations result in an elastic toe-in angle change $\Delta\delta_{vT}$.

With respect to the "power-off effect" in cornering, this will be welcomed at a front wheel, since the outer front wheel will generate an understeering toe-out angle with released accelerator pedal.

On a rear wheel, however, toe-out under traction force is desirable to achieve an understeering toe-in effect with power change at least at the outer wheel. The option of installing an angle α between the two lower transverse links has already been used up by the braking-force layout, and it will normally be improbable that the same angle is suited to generate the necessary correcting angle $\Delta\delta_{Tcorr}$. There is, however, an ultimate measure left, namely the mutual adaptation of the effective levers b_B and b_T in such a manner that both cases of load can be solved by using the same angle α of inclination [11], see Figures 21.54a and c.

On a steerable suspension, of course, the levers b_B and b_T correspond to the scrub radius r_S and the wheel-center offset r_C, and the elasto-kinematic requirements may clearly become conflicting with those of steering geometry.

If there is a hub reduction gear, the wheel carrier is loaded—in addition to the force F_T at the wheel center—by the clockwise reaction torque of the gear $F_TR(i - 1)/i$ with the tire radius R and the reduction ratio i.

Nonetheless, the preceding considerations should have made it plausible that the elasto-kinematic harmonization of a wheel suspension can—at least theoretically—be done covering all of the relevant cases of external load, like wheel load, lateral force, braking, and possible traction force, by suitable choice of the spring rates of the suspension joints and of the positions and inclinations of the links (clearly, this in reality would, of course, require that the investigations be iterated). On the other hand,

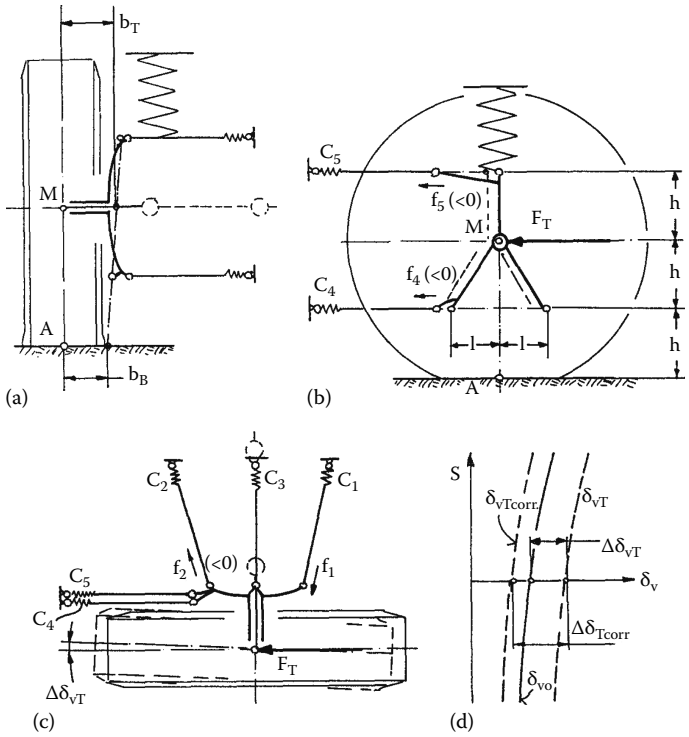


FIGURE 21.54 Suspension of Figure 21.51 loaded with a traction force F_T.

it may have shown up that the more promising elasto-kinematic harmonization is, the more suspension links are available. This is the main reason for the increasing application of multi-link suspensions.

A driven five-link rear axle of a production car, which reveals application of all of the measures discussed earlier, is depicted in Figure 21.55. Upper and lower diagonal links provide lateral and longitudinal guidance of the wheel carrier, also supporting the braking torque (of course with elastic "windup"). All the links have different angles of inclination in space. The elongated axes of the two lower links intersect outside their outboard joints while the two upper ones cross. In contrast to the right-hand side of Figure 21.33, this leads to a "virtual axis" i that is inclined upward—inside in the vehicle—clearly a measure of harmonization of braking and traction-force levers, as discussed with Figure 21.54. Such a position of the axis i could, of course, also have been achieved by using an upper triangular link; however, this (shorter) transverse link would reduce the roll-center height change with wheel travel (see Figure 21.28) and reinforce the disadvantageous jacking-up effect in cornering.

The simple suspension of Figures 21.51 through 21.54 with its parallel longitudinal links will not change its elasto-kinematic properties seriously with wheel travel. However, many suspensions—especially rear ones—show a rotation of the wheel carrier with wheel travel in side view. This is typical for a system with noticeable braking and traction-force support angles (Figure 21.56). If designed for a "neutral" reaction to lateral forces, by locating the "neutral" point of attack of forces at a distance of the pneumatic trail n_R behind the tire contact point A, this neutral point (which is "fixed" to the wheel carrier!) will move along a path g_N about the pole L with wheel travel. The point of attack of the lateral force F, however, moves along a path g_F that is parallel to the path g_M of the wheel center M. In cornering, with the outer wheel suspension compressed and the inner released, consequently with the lever of the outer force increased and that of the inner reduced, both lateral forces will generate considerable elastic toe-in angles (of course, tolerable at rear wheels).

Rear suspensions are frequently mounted on subframes that are connected to the body by bulky rubber mountings for good noise insulation. Moreover, with simple suspension systems that are not

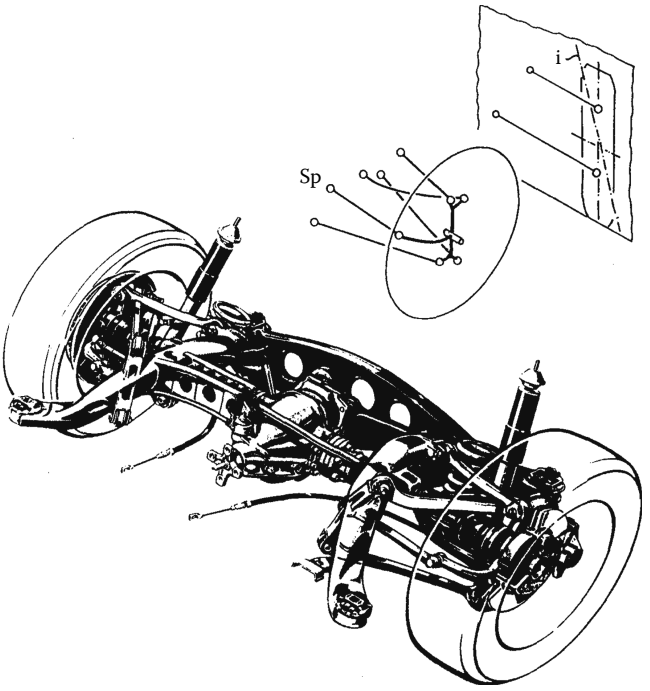


FIGURE 21.55 Five-link rear suspension for optimized elasto-kinematic harmonization. (Courtesy of Daimler Chrysler AG, Stuttgart, Germany, 1991.)

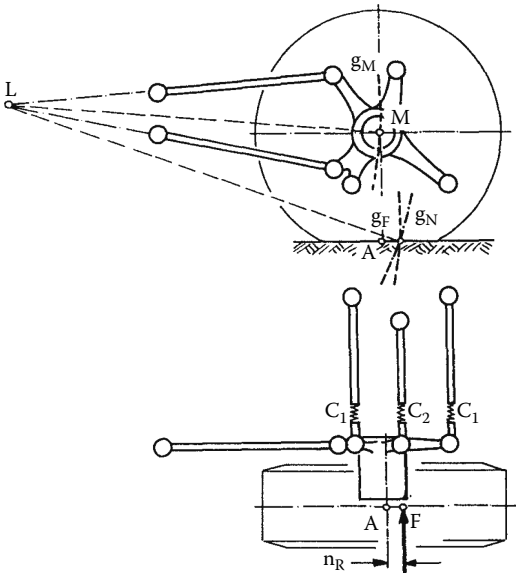


FIGURE 21.56 Variable elasto-kinematic properties caused by a rotation of the wheel carrier in side view.

suited for a deliberate elasto-kinematic harmonization, a subframe may be the only possibility to achieve elasto-kinematic effects.

Since the mountings of a subframe are normally positioned symmetrically to the medial plane of the vehicle (Figure 21.57a), the "elastic center" S_F of the system lies in the medial plane. Unilateral fore-and-aft forces therefore cause elastic steering angles of the whole assembly. On a rear axle, a unilateral braking as well as traction force generates a steering angle that counteracts the moment

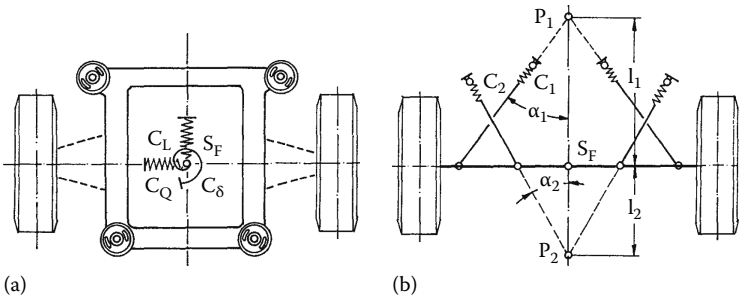


FIGURE 21.57 Subframe with elastic mountings (a) and rigid-axle suspension; (b): c = Spring rates, C_δ = angular spring rate; and S_f = elastic center.

of the unilateral force about the vehicle's center of gravity and so can be useful. On a front axle, of course, a subframe leads to a steering angle that reinforces the yaw reaction of the vehicle to the unilateral force. This is why subframes are rare on front suspensions, except for aggregates with wide-based arrangement of their mountings for maximum rotational stiffness about the vertical axis.

A rigid axle with its suspension links and their elastic joints (Figure 21.57b) shows the same basic properties as an elastically mounted subframe.

A subframe normally represents a highly overconstrained structure, the deformations of which depend on the combination of all external forces acting together. It is therefore impossible to determine a "spring rate" of a mounting point of a suspension link, with the aim to use this similarly to the spring rate of a rubber joint.

By skillful design, however, the influence of the deformations of a subframe on sensitive suspension properties can be minimized. Figure 21.58 shows a driven rear axle, its suspension being a derivative of the trapezoidal-link system fairly resistant to braking windup (see also Figure 21.9), however with two upper transverse rod links for better toe-in control. The subframe is cranked to let the propeller shaft and the exhaust pipes (not shown) pass through and, consequently, is relatively "weak" in its lower region.

Under traction torque, the forces of the front joints of the lower trapezoidal links act toward the inner side of the vehicle, and the force of the front mounting of the gearbox acts in vertical direction, both tending to bend the front transverse beam of the subframe upward. The contrary happens at the rear transverse beam. Consequently, the longitudinal beams of the subframe are subjected to a

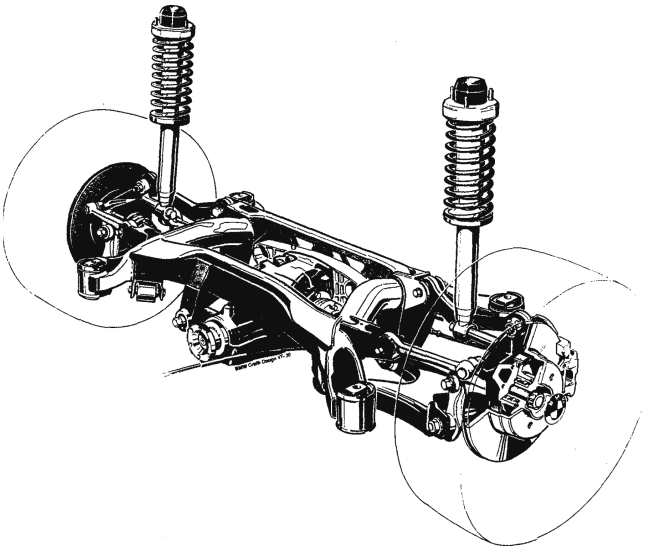


FIGURE 21.58 Multi-link rear suspension. (Courtesy of BMW AG, Munich, Germany, 1994.)

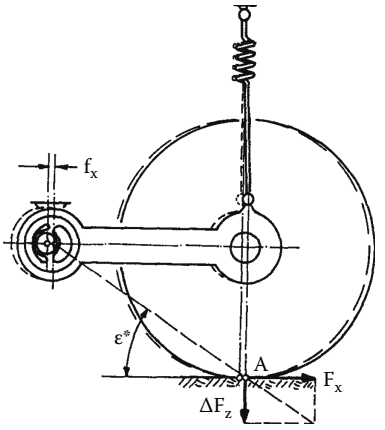


FIGURE 21.59 Compliant suspension under braking force.

twisting torque. Horizontal transverse displacements of the inner joints of the two upper rod links and, with that, undesired steering angles are avoidable by positioning the links at the height of the longitudinal beams of the subframe—which forms an almost planar rectangle in this region and therefore provides sufficient stiffness in any horizontal direction.

Finally, a question may be discussed briefly that often arises among suspension engineers: whether there is an influence of the suspension's compliance on the kinematic suspension characteristics or not.

The trailing-arm rear suspension in Figure 21.59 has a rubber joint for good longitudinal compliance and is here loaded by a braking force F_x combined with the respective "dynamic" wheel-load release ΔF_z. Since the resulting force of F_x and F_z acts in the direction of the joint, the state of "100% anti-lift" geometry is achieved. However, the braking force causes an elastic rearward deflection f_x. This deflection obviously does not lead to any change worth mentioning regarding the static diagram of forces. That, which is clearly to be seen with this simple mechanism, also applies to more complicated suspensions: the elastic deflection of a (statically determinate) mechanism is of negligible influence on the positions of the links and so on the static diagram of forces.

Any attempt to define "suspension characteristics" that consider the elastic deformations will be useless and misleading. For instance, an elastic longitudinal displacement of 5 mm per 100 mm wheel travel would simulate a support-angle contribution of 3°, meaning about 15%–20% anti-dive geometry at a passenger car front axle or about 25% anti-lift or anti-squat geometry on a vehicle with single-axle drive. For the same reason, attempts to determine suspension characteristics by measuring the path of the wheel on a realistic system will give dubious results concerning the internal reaction forces and torques of the system with wheel travel. The only reliable method for determining kinematical suspension characteristics is to consider the (theoretical) "stiff" mechanism—which, of course, means that they best follow from calculation of the kinematic model.

21.8 CONCLUSION

A wheel suspension is a kinematic mechanism that enables deliberate influencing of the vehicle's reactions to vehicle-dynamic events like traction, braking, and cornering.

An attempt has been made in this treatise to touch upon the most important aspects of and influences on suspension design, including the effects of such elements as driveshafts and hub reduction gears and of elastic mountings.

For more precise information about related topics, special books [12] and standard works or handbooks are available on vehicle dynamics, tire properties, and vibrations [13–19] as well as construction elements like brakes [20], driveshafts [2,21], metallic springs [22,23], or rubber springs [24,25].

NOTATIONS

(In the text, vectors are characterized by **bold** letters.)

A	Tire contact point
c	Spring rate (N/mm)
d	Kingpin axis; axis of rotation
F	Degree of freedom of a mechanism
F	Force (N)
f	Degree of freedom of a joint
f	Spring deflection (mm)
h	Center-of-gravity height (mm)
h_{RZ}	Roll-center height (mm)
i	Transmission ratio
K	Wheel carrier
l	Wheelbase; length (Mm)
M	Wheel center
M	Moment, torque (N mm)
m	Instantaneous axis
m	Mass, inertia (kg)
n	Castor offset (mm)
n_R	Pneumatic trail (mm)
n_τ	Castor offset at wheel center (mm)
P	Pole
p	Polar distance (mm)
p	Wheel-load lever arm (mm)
R	Tire radius (mm)
RZ	Roll center
r	Roll axis of the vehicle
r_C	Wheel-center offset (mm)
r_S	Scrub radius (kingpin offset) (mm)
r_T	Traction-force radius (mm)
S, SP	Center of gravity
s	Instantaneous-screw axis
s	Wheel travel (mm)
t	Time (S)
v	Velocity (mm/s)
x, y, z	Axes of the coordinate system
α	Deflection angle of a driveshaft joint (rad)
α	Tire sideslip angle (rad)
γ	Camber angle (rad)
δ	Steering angle (rad)
δ_v	Toe-in angle (rad)
ε	Wheel-travel angle (rad)
ε^*	Support angle (hub-mounted torque support) (rad)
ε^{**}	Support angle (chassis-mounted torque support) (rad)
ϑ	Pitch angle (rad)
σ	Kingpin inclination (rad)
τ	Castor angle (rad)
φ	Roll angle (rad)
$\boldsymbol{\omega}$	Angular velocity (rad/s)

REFERENCES

1. Matschinsky, W.: *Radführungen der Straßenfahrzeuge*. Springer, Berlin, Germany, 2007.
2. Seherr-Thoss, H.-Ch. Graf von, Schmelz, F., and Aucktor, E.: *Universal Joints and Driveshafts*. Springer, Berlin, Germany, 1992.
3. Behles, F.: Die Beherrschung des Brems- und Anfahrnickens. *Automobiltechnische Zeitschrift* 66 (1964) 225–228.
4. Winkelmann, O. J.: Anforderungen an das Fahrverhalten von Kraftfahrzeugen. *Automobiltechnische Zeitschrift* 63 (1961) 121–128.
5. Matschinsky, W.: Der Einflurß der Antriebsart auf die Radführung—besonders bei Zweiradfahrzeugen. *Automobiltechnische Zeitschrift* 67 (1965) 221–225.
6. Matschinsky, W.: *Road Vehicle Suspensions* (Translation edited by A. Baker). Professional Engineering Publishing, London, U.K., 2000.
7. Matschinsky, W., Dietrich, C., and Winkler, E.: Die Doppelgelenk-Federbein-Vorderachse der neuen BMW Sechszylinderwagen der Baureihe 7. *Automobiltechnische Zeitschrift* 79 (1977) 357–365.
8. Dingerkus, O.: Über die Spur- und Sturzänderung spezieller Radaufhägungen von Kraftfahrzeugen. *Automobiltechnische Zeitschrift* 65 (1963) 49–55.
9. Forkel, D.: *Ein Beitrag zur Auslegung von Fahrzeuglenkungen*. Deutsche Kraftfahrzeugforschung und Strarßenverkehrstechnik No 145. VDI-Verlag, Düsseldorf, Germany, 1961.
10. Behles, F.: Möglichkeiten und Grenzen der Verbesserung der Federweichheit von Kraftfahrzeugen. Diss. München, Germany, 1962.
11. Braess, H.-H. and Ruf, G.: Influence of tire properties and rear axle compliance steer on power-off effect in cornering. *Proceedings of the 6th International Conference of Experimental Safety Vehicles*. Washington, DC, 1976.
12. Bastow, D.: *Car Suspension and Handling*, 2nd edn. Pentech Press, London, U.K., 1988.
13. Fenton, D.: *Handbook of Vehicle Design Analysis*. Professional Engineering Publishing, London, U.K., 1996.
14. Fenton, D.: *Handbook of Automotive Powertrain and Chassis Design*. Professional Engineering Publishing, London, U.K., 1998.
15. Ellis, J.R.: *Vehicle Handling Dynamics*. Professional Engineering Publishing, London, U.K., 1994.
16. Mitschke, M.: *Dynamik der Kraftfahrzeuge* Vol. A: *Antrieb und Bremsung*, 3rd edn., 1995 Vol. B: *Schwingungen*, 3rd edn., 1997 Vol. C: *Fahrverhalten*, 2nd edn., 1990, Springer, Berlin, Germany.
17. Braess, H.-H. and Seiffert, U. (eds.): *Handbuch Kraftfahrzeugtechnik, Vieweg*, Springer, Berlin, Germany, 2000.
18. Fenton, D.: *Advances in Vehicle Design*. Professional Engineering Publishing, London, U.K., 1999.
19. Matthews, C.: *I Mech E Engineers' Data Book*, 2nd edn. Professional Engineering Publishing, London, U.K., 2000.
20. Harper, G.A.: *Brakes and Friction Materials*. Professional Engineering Publishing, London, U.K., 1997.
21. Seherr-Thoss, H.-Ch. Graf von, Schmelz, F., and Aucktor, E.: *Gelenke und Gelenkwellen*. Springer, Berlin, Germany, 2002.
22. Gross, S.: *Berechnung und Gestaltung von Metallfedern*. Springer, Berlin, Germany, 1960.
23. Gross, S.: *Calculation and Design of Metal Springs*. Chapman & Hall, London, U.K., 1966.
24. Göbel, E.F.: *Berechnung und Gestaltung von Gummifedern*. Springer, Berlin, Germany, 1955.
25. Göbel, E.F.: *Rubber Springs Design*. Butterworth, London, U.K., 1974.

22 Active and Semiactive Suspension Systems

Davor Hrovat, Eric H. Tseng, Michael Fodor, and Jahan Asgari

CONTENTS

22.1 Introduction 769
22.2 Actuator Hardware Types 771
22.2.1 Electrohydraulic Actuators 772
22.2.2 Electrical Actuators 773
22.2.3 Semiactive Actuators 774
22.2.4 Load Levelers 775
22.3 Hardware-Related Considerations 776
22.3.1 Friction 776
22.3.2 Leakage 776
22.3.3 Actuator Controls: Force versus Displacement Tracking 777
22.3.4 Actuator Bandwidth 779
22.4 Hardware Implementations 779
22.4.1 Narrow-Bandwidth Active Suspension 779
22.4.1.1 Mercedes "Active Body Control" 781
22.4.2 Broad-Bandwidth Active Suspensions 782
22.4.3 Semiactive Suspensions 784
22.5 Vehicle Models 784
22.6 Other Concepts 788
22.6.1 Anti-Dive, Anti-Squat, and Anti-Pitch Suspension Designs 788
22.6.2 Delft Suspension 789
22.6.3 Roll-Control Suspensions 789
22.6.4 Dynamic Vibration Absorbers or Tuned Mass Dampers 791
22.6.5 Additional Considerations 792
22.A Appendix 792
References 794

22.1 INTRODUCTION

One approach to classifying automotive suspensions places them into three categories: passive, semiactive (SA), and active. Passive suspensions are defined here as those that do not require any external power or energy sources; their component tuning is fixed and not adjusted during operation. A typical passive suspension consists of mechanical springs and shock absorbers with no electronic control. SA suspensions contain elements whose behavior is adjusted through computers or similar means but that do not require a significant level of power input for their operation. One example is the adjustable damper whose valving requires some signal level power (typically electrical) but substantially less than a high-force hydraulic piston, for instance. A typical SA suspension is visually similar to a passive system but employs electronically controlled dampers with associated sensors (heave, pitch, roll), microprocessor, and wiring. These actuators absorb and dissipate mechanical energy but do not supply mechanical energy to the

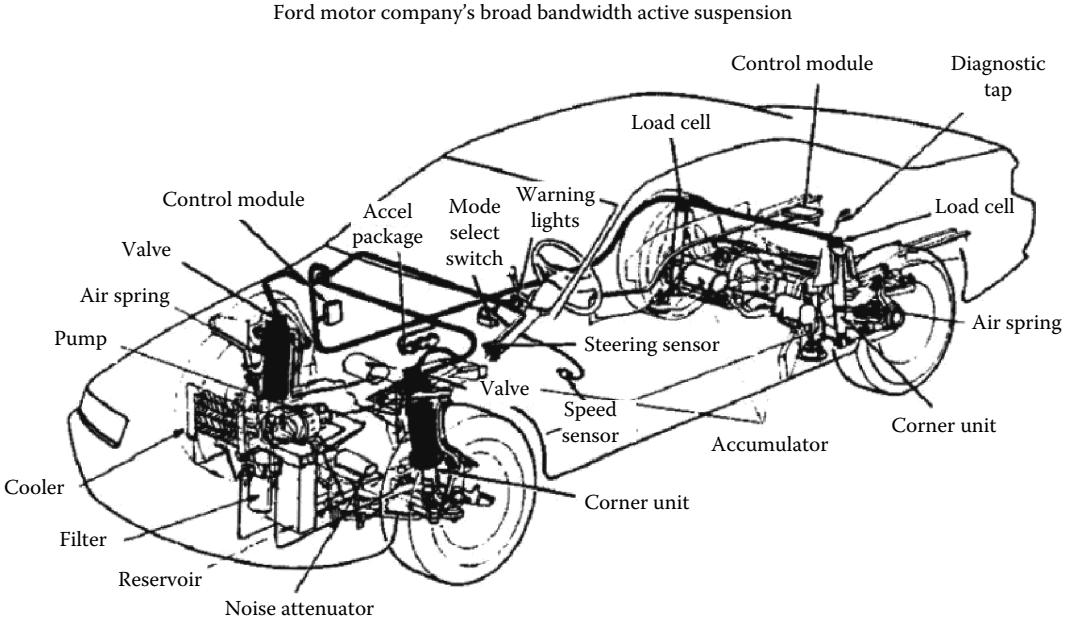


FIGURE 22.1 Typical hydraulic fully active suspension system and components.

rest of the vehicle system. Fully active suspensions contain actuators that are typically controlled through closed-loop means and require significant power for operation. Active systems both dissipate and supply mechanical power to the vehicle system as a whole. System complexity increases further with a typical fully active suspension, which includes the aforementioned electronics, some sort of electrohydraulic actuation with pumps, valves, and associated sensors. Figure 22.1 shows the components of a typical fully active system [1] including actuators, sensors, microcontroller, and hydraulic supply system.

Chapter 31 describes the overall system design concepts and goals for active suspension design. These require simultaneously managing the trade-offs between ride comfort, road holding, and allowed suspension stroke or rattlespace. These theoretical results provide indicators of potential active suspension benefits and corresponding requirements for actual hardware development and implementation including actuator force capability, power, and bandwidth. Other concerns include packaging, cost, reliability, serviceability, noise/vibration/harshness (NVH), and fault tolerance. The various control approaches described in Chapter 31—skyhook control, optimal control, etc.—produce desired actuator forces that can be considered as part of an outer loop. These in turn provide the command or reference values for the inner-loop force production and tracking as shown in Figure 22.2. The actual delivery of this desired force tracking is a very challenging task in practice (see Section 22.3). Some of the challenges include high-fidelity force control in the face of the always present disturbances due to road irregularities, limited actuator bandwidth, stroke and force authority, and inherent system friction.

The payoff for a well-designed system can be high given the potential system-wide benefits to vehicle dynamic performance—ride, handling, safety, etc. Driver reactions to such a system can be quite positive. Active suspensions attracted considerable attention in the mid-eighties with the widely publicized demonstrations of the Lotus Esprit experimental vehicle. This car evolved through many years of R&D work. A January 21, 1987, *The New York Times* article describes the vehicle capabilities as follows:

> Test drivers usually emerge from the car with their imagination in overdrive. 'The greatest single advance in car engineering since the war,' the British magazine *Car* declared on the cover of a recent issue. Car's editor, Steve Cropley, wrote that one could take the benefits of all other modern automobile developments, 'add them up and double the total—and you might come somewhere near the degree to which full active suspension improves a car.'

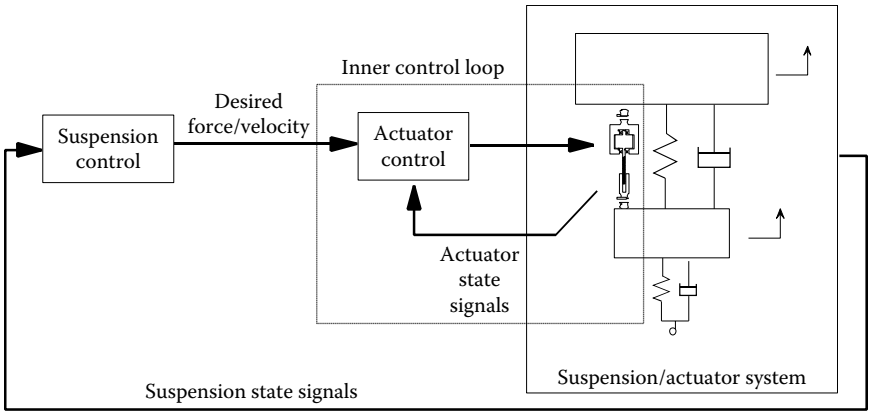


FIGURE 22.2 Inner-/outer-loop suspension control.

Doug Nye, a contributing editor of the American magazine *Road & Track*, said he was initially skeptical, but 'in fact it is amazing'. He said, 'I've been writing about cars for 24 years and honestly found it the most impressive thing I've ever tried'.

As always, these statements contained some hype and should have been taken with a grain of salt. In particular, the earlier comments refer mainly to substantial improvements in vehicle handling under somewhat specialized and severe maneuvers. The corresponding improvement in ride comfort, which is probably more relevant to an average consumer, was not clearly demonstrated then and was the subject of subsequent work. Nevertheless, the early demonstrations pointed to potentially "leap-frog" function improvements in vehicle ride and handling, with related impacts on overall vehicle controls and safety. This, in turn, spurred increased industry-wide activities, which eventually resulted in the introduction of the first narrow-bandwidth or "slow-acting" active production units from Nissan and Toyota in the 1990s and subsequent entries from Citroen (active roll control) and Mercedes (the so-called "Active Body Control" (ABC) suspension described in Section 22.4.1.1).

The following sections focus on the hardware side of active and SA suspension design including implications for control. Section 22.2 presents various types of active and SA actuator hardware. Section 22.3 describes hardware-related considerations related to control system design. Section 22.4 discusses various hardware implementations of advanced suspension concepts. Relevant vehicle models are presented in Section 22.5. Related considerations are summarized in Section 22.6.

22.2 ACTUATOR HARDWARE TYPES

Implementation of active and SA suspensions has been accomplished with a wide variety of hardware, each with its own design trade-offs. These trade-offs include typical engineering concerns such as performance, cost, complexity, and reliability. Performance generally reduces to the suspension actuator's bandwidth limit or its ability to reject roadway disturbances of increasingly higher frequency. As a result, fully active suspensions can be divided into two groups—high (or wide or broad) and low (or narrow) bandwidth. High-bandwidth actuators are defined here as those capable of controlling the higher-frequency wheel-hop suspension mode (roughly 10 Hz) as well as the lower-frequency body mode (roughly 1 Hz). These require force control capability well above 10 Hz. Narrow- or low-bandwidth actuators control only the body mode and can relax the bandwidth requirement to around 5 Hz; typically they are characterized by reduced power requirement. However, by definition they would also imply inferior performance as compared to the high-bandwidth systems. Figure 22.3 shows typical topologies for high- and low-bandwidth

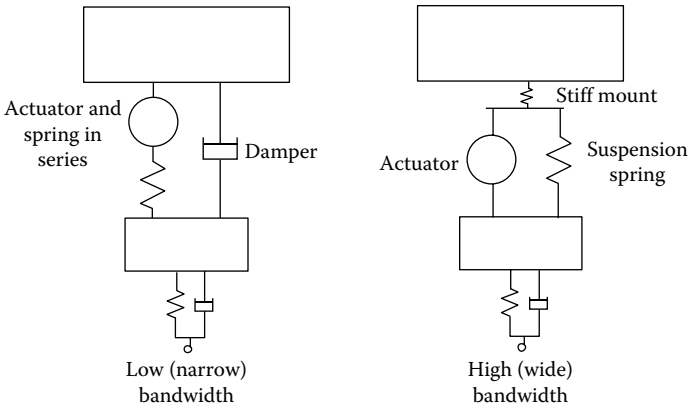


FIGURE 22.3 Typical high- and low-bandwidth suspension topologies.

systems. To fully exploit the high-bandwidth force capability of an actuator, a designer should avoid introducing too much force-filtering compliance between the actuator and the sprung and unsprung masses. Thus, traditionally, high-bandwidth systems contain only a relatively stiff mount and operate in parallel with the primary suspension compliance. The primary suspension compliance in a low-bandwidth system is typically a coil spring or some accumulator of the same type as the actuator (i.e., hydraulic/pneumatic) placed in series with the active actuator. SA suspension actuators are usually dampers with adjustable rate. They too are subject to bandwidth constraints, which limit their potential effectiveness. The following section presents electrohydraulic, electric, SA, and load-leveling actuators and some of their associated analysis and design considerations.

22.2.1 Electrohydraulic Actuators

Electrohydraulic actuation is by far the most common seen in experimental and production fully active suspensions. Hydraulic piston–cylinder units offer the necessary force capability and bandwidth for active suspension control while meeting typical suspension packaging requirements. Drawbacks include inherent energy losses due to perpetual fluid pumping and the service and maintenance requirements of a system of couplers and hoses filled with hydraulic oil. Figure 22.4 shows simplified schematic diagrams for both low- and high-bandwidth hydraulic actuation methods. For both systems, a servo valve is used to move high-pressure fluid into and out of a cylinder. The low-bandwidth actuator contains a large hydraulic compliance in series with the piston in the form of a large accumulator. This tuned element "filters" the hydraulic fluid pressure fluctuations acting on

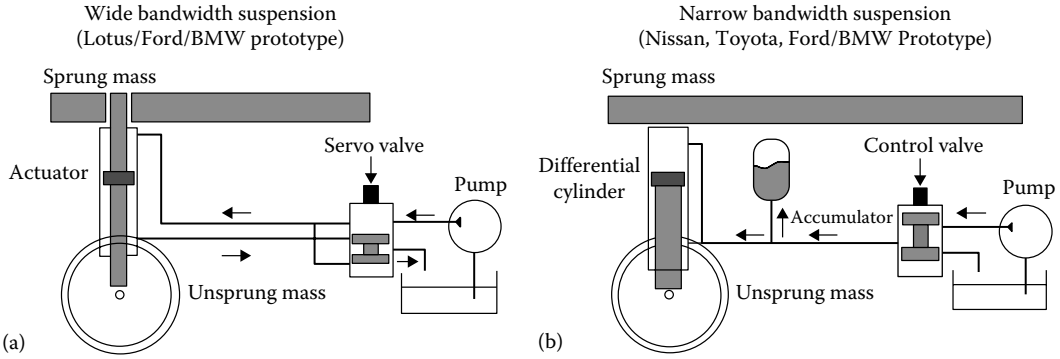


FIGURE 22.4 Typical high-/low-bandwidth (a)/(b) electrohydraulic suspension layouts.

the accumulator/piston pair, thereby reducing the effective actuation bandwidth of the system while at the same time diminishing the effects of (higher-frequency) road input disturbances.

Typical electrohydraulic actuator model is summarized in Appendix 22.A. In the same appendix, this actuator model is applied to a simple quarter-car model of vehicle ride dynamics.

22.2.2 Electrical Actuators

Electrical actuation may offer several possible advantages including high-force production bandwidth, low NVH, ease of serviceability, and the opportunity for energy regeneration. These advantages are potentially offset by issues relating to weight and limited maximum force capability. High-bandwidth systems may also require higher-voltage electrical infrastructure in the vehicle—a potentially costly requirement. A potentially attractive, electromechanical implementation of a practical broad-bandwidth actuator has been proposed in [2,3]. The proposed electrical active suspension (EAS) actuator, shown in Figure 22.5, consists of a ball–nut combination, which is used to transform the rotational power from a highly efficient permanent magnet motor into a corresponding suspension force and translational velocity. The stator, which carries the majority of the actuator's mass, is mounted to the sprung mass (body) of the vehicle leaving only the ball nut and lower bushing/mount to contribute to the unsprung (wheel) mass. In addition to facilitating higher bandwidth, the electronic controls for the unit allow energy regeneration when the actuator is removing mechanical energy from the suspension system. The primary challenge controlling the actuator unit arises from the inertial coupling of the sprung and unsprung masses by the rotating inertia of the actuator. Any motion across the sprung and unsprung masses results in rotation of the motor's armature. This inertial motion must be carefully controlled to guarantee that the actuator does not act as a rigid member connecting the vehicle body and wheel at high frequencies. The same phenomenon drives the design of the upper and lower actuator bushings to ensure a reasonable compliance for the whole device at high frequency. Additional details can be found in [4].

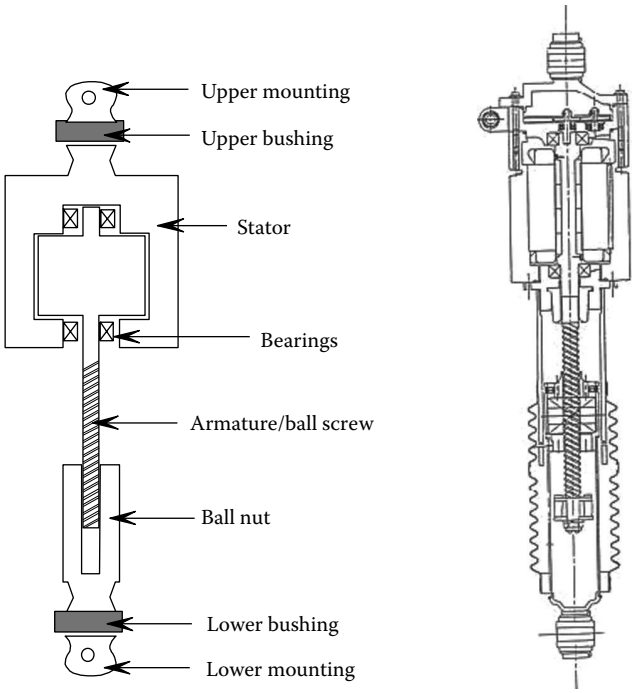


FIGURE 22.5 Ford electric active suspension actuator concept.

While the earlier concept is based on a rotational electromechanical motor or power source, a more recent approach based on the linear motor motion and permanent magnets has been mentioned in [5]. In this case, the very effective control of suspension forces is claimed based on the experience gained from many years of successful production of premier speaker systems from the Bose Corporation. According to [5], the suspension system also contains the so-called damper weight that helps to reduce the wheel hop, which is reminiscent of the dynamic absorber (DA) or tuned mass damper discussed in Section 22.6.4. According to [5], the linear motion suspension concept seems promising; however, no actual vehicle test data have been reported thus far and it will be interesting to see future progress with this intriguing concept.

22.2.3 Semiactive Actuators

SA control is usually accomplished using an adjustable damper in parallel with the main suspension spring. Damping rate for each actuator is determined by the chosen control law as described in Chapter 31. Adjustable dampers can be constructed from pneumatic or hydraulic piston–cylinder combinations with electromechanical control of an orifice usually placed in the piston face. Actuator bandwidth is then established by the reaction time of the controlling valve and associated force production dynamics. Alternatively, magneto-rheological (MR) fluids can be employed to change the viscosity of the working fluid, which, in turn, changes the force across the actuator. This method benefits from faster response time, although limited fluid life can contribute to service concerns. One MR damper application is found in the 2002 Cadillac Seville STS where its MR fluid, codeveloped by Delphi and Lord Corporation, is composed of suspended iron particles in a base fluid of synthetic hydrocarbon. Its continuous variable damping is adjusted at 1000 Hz. The fast response time of the MR damper enables closed-loop vehicle stability control through the control of lateral and longitudinal load transfer characteristics of the suspension during transient movements [6].

One method for reducing the bandwidth requirement of the orifice control valve is to employ the resistance control SA damping method demonstrated experimentally in [7]. Here, a dual-acting damper is used that contains two control valves, each setting the damping rate of the device in only one direction of operation (jounce or rebound). The orifices are switched in and out using passive check valves. These passive valves provide the high-bandwidth switching required by the high-frequency tire motion, while the low-bandwidth control valves are only required to move in response to the relatively low-bandwidth body motion. Figure 22.6 demonstrates the operation of this device. Depicted are sprung mass velocity, $\dot{z}_s$; relative velocity across the damper, $\dot{z}_s - \dot{z}_{us}$; the controlled damping rates used for jounce and rebound dampers (as a function of the body motion); and the resulting combined damping rate for the device, b_s. The combined damping rate provides the high-frequency changes desired while using only low-bandwidth control valves. A cost savings is potentially realized by replacing a single high-performance, high-bandwidth valve with two lesser performing valves of lower total cost.

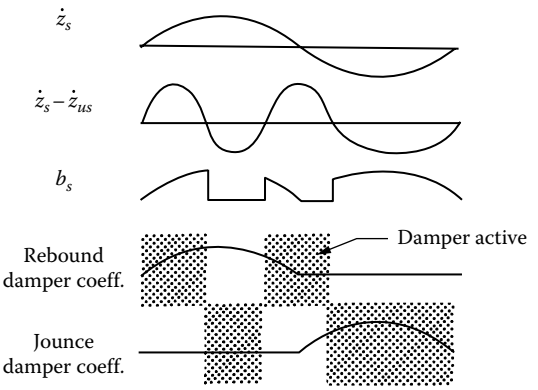


FIGURE 22.6 Operation of combined dampers.

22.2.4 Load Levelers

Load-leveling suspensions are constructed topologically similar to low-bandwidth suspensions in that the equilibrium point of the main suspension spring is the primary adjustment. However, they operate in a very low-bandwidth manner, being incapable of controlling even the low-frequency body mode. Instead, they perform similarly to passive suspensions but offer the capability of "rebalancing" the posture (or attitude) of the vehicle to maintain an appropriate vehicle stance under a wide range of loading conditions and properly center the suspension in the available rattlespace. As such, they act as a slow integrator. They also allow for a softer main spring and are commonly seen on luxury vehicles with a plush ride.

Load-leveling suspensions are often constructed using pneumatic springs whose equilibrium length is adjusted by changing the volume of air in the spring. This is accomplished by means of an air pump and simple on/off valve. Bang–bang control is typically used to set the spring length according to a set of ride height sensors. Typical examples of electronically controlled load leveling can be seen in 1984 Lincoln Continental. The electronically controlled air suspension system (see Figure 22.7a and b) allows its spring to be 32%–38% softer than a corresponding steel spring suspension while maintaining vehicle stance and suspension travel [8]. In this and most other load-leveling suspensions, height sensors are used to trigger the adjustment of the air amount in the spring. In Lincoln Continental, signals from the ignition, doors, and brake systems are also used to ensure leveling only when necessary. For example,

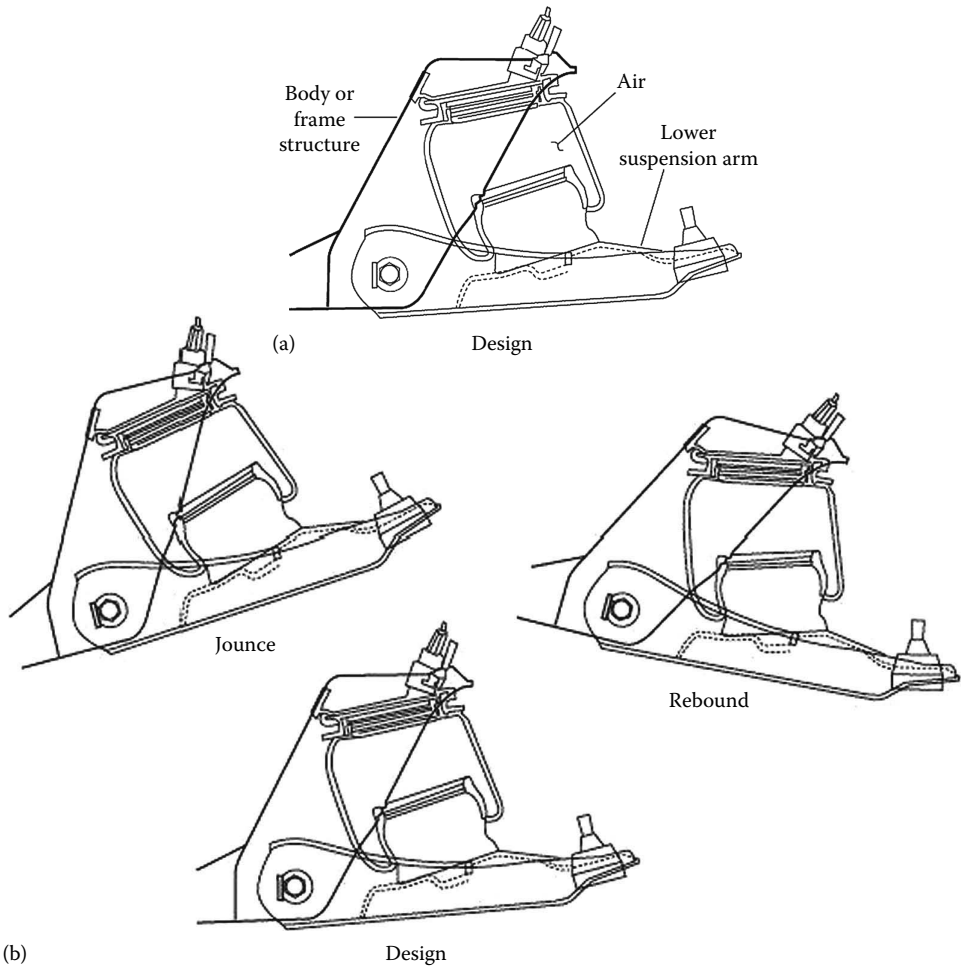


FIGURE 22.7 (a) Typical spring-on-arm air spring and (b) air spring operational modes.

if the ignition is on and the door is opened, the vehicle will level within 50 s to account for additional driver and passenger weight. If the ignition is off, only limited leveling events will occur.

22.3 HARDWARE-RELATED CONSIDERATIONS

Active suspension control system design is challenging in its requirement to balance several design trade-offs such as ride comfort, handling, and rattlespace use simultaneously. Often, the design analysis idealizes the actuators, assuming that they operate at limitless bandwidth or that their bandwidth characteristics are uniform over the entire range of operation conditions. These simplifications, while helpful for system-level analysis, need to be reconsidered when pursuing practical suspension realizations. Thus, the control problem is further complicated when more detailed hardware considerations are included. Typical hardware concerns are actuator bandwidth limits, friction effects in the actuators and associated hardware, and determining the appropriate control system topology for a given actuator. These are discussed in more detail in the following text.

22.3.1 Friction

Even a relatively small amount of dry friction can substantially degrade ride performance and lead to a significant discrepancy between the projected and actual ride benefits achieved with active and SA suspensions. For example, for a 3000 lb/1364 kg vehicle with a per-corner weight of 750 lb/341 kg, a corresponding actuator dry friction of only 25 lbf/111 N can contribute to a respectable 3.3% g root mean square (rms) sprung-mass acceleration. To see this, assume that the desired acceleration (or corresponding net sprung-mass force) is zero, that the actuator bandwidth is well below the wheel-hop frequency, and that the wheel-hop dynamic motion is sufficiently agile to "break" the friction (if it were not, then the suspension would lock and result in the undesirable situation of "ride-on tires" where both vehicle sprung and unsprung masses would ride on tire spring).

22.3.2 Leakage

There are a number of possible leakage paths for an electrohydraulic active suspension actuator. This includes the leakage across the piston seals, leakage across the supply and vent ports, as well as cross-port leakage. In particular, the cross-port leakage can be a significant factor influencing the overall system dynamic behavior [9]. In effect, this leakage acts in series with the hydraulic oil compliance; this can be seen by imagining the actuator servo valve operating at/around its centered position (see Figure 22.8). The net effect of this is that the corresponding vehicle/suspension modes have more damping, especially in the dominant, sprung-mass-related mode.

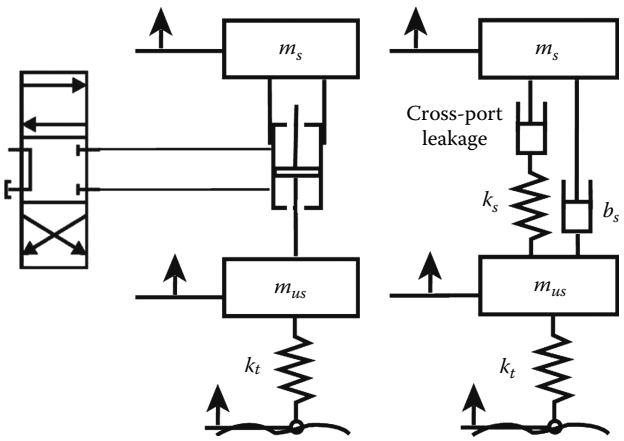


FIGURE 22.8 A wide-bandwidth active suspension schematic and corresponding model.

22.3.3 Actuator Controls: Force versus Displacement Tracking

An important aspect of practical significance for active suspensions is the issue of actuator dynamics and its interaction with vehicle system dynamics. In practice, the challenging task of high-fidelity actuation requires a good understanding of underlying system dynamics and a mastery of related mechatronic design skills. For example, consider the earlier hypothetical case of the constant, for example, zero net desired sprung-mass acceleration for the best possible ride comfort. Even this seemingly inconspicuous and easy task of producing the constant actuator force equal to the sprung-mass weight can be quite challenging. Chiefly, as outlined in [10], the main issue here is the suspension force variations created by relative velocity motion across the suspension mounting points, which in effect acts as a force disturbance generator. Elimination of these effects is one of the main tasks of the inner-loop force delivery subsystem.

Note that the elimination or containment of these disturbances also precludes or diminishes the introduction of any additional vehicle dynamics into the inner-loop force-tracking subsystem. It has been shown in [11] that through this additional dynamic effect, the poles of the base (vehicle) plant become zeros of the resulting closed-loop system. More precisely, the inner-loop force-tracking system can be represented via a block diagram shown in Figure 22.9, where C_{in} represents the inner-loop controller and G_a and G_p are the actuator and plant transfer function, respectively. The corresponding transfer function between the desired and actual force delivered at the suspension actuator is

$$G_{ff} = C_{in}\frac{G_{fi}}{1 + C_{in}G_{fi}} \tag{22.1}$$

where

$$G_{fi} = \frac{G_a}{1 + G_pG_h} \tag{22.2}$$

Here, G_h summarizes the feedback path created through actuator hardware, which represents suspension force changes due to relative motion across the suspension mounting points. This coupling or feedback effect is present with most hardware implementations, although the degree of the coupling may vary and depends on the kind of suspension implementation concept one is dealing with (electrohydraulic vs. electrical). In general, it is more difficult to exactly counteract this term in the case of electrohydraulic actuators due to many highly nonlinear effects—such as viscosity temperature dependence and air entrainment—and some form of adaptation will most likely be needed, as elaborated in [12]. For completeness, Figure 22.9 also shows an outer or main loop controller C_{out}, which may include an appropriate observer or (Kalman) filter.

From the earlier transfer functions (Equations 22.1 and 22.2), one can see that the open-loop poles of the vehicle plant (G_p) become the zeros of the resulting force-tracking inner-loop configuration (G_{ff}). In the case of a typical quarter-car vehicle model, this means that there are two pairs of relatively lightly damped zeros in this inner loop. The first pair corresponds to the dominant, sprung-mass-related heave mode of vibrations, which is typically of relatively low frequency

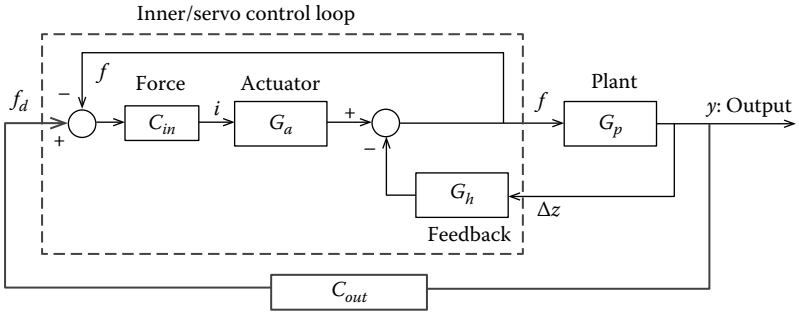


FIGURE 22.9 Force-tracking actuator controller design.

around 1 Hz. The amount of damping in this mode may vary, but it can be close to the corresponding passive case where the damping ratio is around 0.2–0.3. For example, incorporating a passive or SA damping component as an integral part (or in addition) to an active suspension actuator, one can relatively easily increase the damping ratio of the earlier dominant mode to 0.1–0.2.

The second pair of zeros corresponds to the wheel-hop mode, which typically resides around 10 Hz with relatively small amount of inherent damping. As such this mode is much more challenging to contain. This pair of zeros will be the limiting factor for high-performance system capabilities, due to the fact that the resulting closed-loop poles will tend toward system zeros. One way to improve this situation and increase the effective wheel-hop damping is through structural hardware changes by introducing the dynamic vibration absorbers tuned at the wheel-hop frequency. This can also improve the overall system ride and handling performance, as discussed in Chapter 31. It is interesting that one of the best-ride small vehicles ever produced, the legendary Citroen 2CV, embodied the dynamic vibration absorbers for the wheel-hop mode. While this was done in the context of an essentially passive suspension, based on Chapter 31 it is anticipated that a well-developed future active suspension system will incorporate some form of passive or (semi)active vibration absorbers.

A different approach to eventually achieving the same or similar ride and handling performance as with the earlier force tracking was proposed in [12,13]. This time instead of following the desired force command, the inner loop is to be designed to follow the (equivalent) desired suspension displacement traces. This configuration can be represented in a block diagram form as shown in Figure 22.10. The corresponding transfer function between the desired and actual displacement delivered at the suspension actuator is

$$G_{zz} = C_{in} \frac{G_{zi}}{1 + C_{in} G_{zi}} \tag{22.3}$$

where

$$G_{zi} = \frac{G_a G_p}{1 + G_p G_h} \tag{22.4}$$

While with this approach the vehicle plant poles are no more the overall system zeros, the resulting set of zeros now contains the plant zeros, which physically corresponds to the case where the sprung and unsprung masses ride together on tire spring. Neglecting the tire damping, the resulting zeros now form a pure complex conjugate pair and as such they are more challenging than the previously

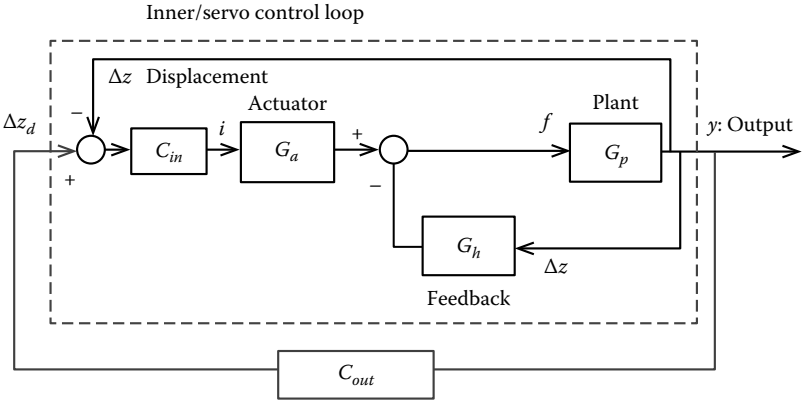


FIGURE 22.10 Displacement-tracking actuator controller design.

discussed sets found with the force-tracking approach. In addition, the limiting frequencies are now smaller, since the frequency of bandwidth limiting zeros from Figure 22.10 is lower than the wheel-hop mode frequency of the bandwidth limiting (lightly damped) zeros from Figure 22.9.

Whether one uses force- or displacement-tracking approach will, among other things, also depend on the prevailing flow or velocities and displacements across the actuator. Large velocities or flows, such as seen with off-road vehicles, may be more effectively dealt with through the displacement-tracking approach. In this case, the flow-control valves can be used too. More about this can be found in Section 22.3.4.

22.3.4 Actuator Bandwidth

When dealing with active suspensions, an important practical consideration refers to actuator bandwidth requirements and limitations. This typically refers to actuator force generation bandwidth for all possible motions across the actuator mounting points. One distinguishes between low- or narrow-bandwidth (or slow-acting) and high- or broad-bandwidth (or fast-acting) suspensions. The former attempts to control only the relatively slow body mode and as such their bandwidth is typically between 3 and 7 Hz. On the other hand, the broad-bandwidth suspensions attempt to control both the body and wheel-hop modes, and thus, their bandwidth should be typically above 12–15 Hz.

One way to control the bandwidth is by introduction of additional compliances in the form of gas–spring accumulators possibly augmented with SA damping. This has a positive side effect of providing a high-frequency, NVH isolation. Additional insight into related structure and dynamics is given in [14], where typical narrow-bandwidth suspension configurations were analyzed with the help of bond graph modeling methodology applied to simple, quarter-car vehicle models shown in Figure 22.11. A generic suspension configuration from Figure 22.11a includes a (possibly variable) hydraulic damping valve, hydropneumatic accumulator, and two oil flow sources, Q_A and Q_B, controlled by appropriate electrohydraulic valves. The equivalent, all mechanical, structure can then be derived with the help of a bond graph from Figure 22.11b with a transformer ratio set to one. The resulting structure is shown in Figure 22.11c. Note that the hydraulic damping and pneumatic compliance are in effect acting geometrically in parallel with each other.

In practice, only one of the two flow sources will typically be present. One example when only V_A is present is shown in Figure 22.11d. This configuration is typical of air suspensions actuators. On the other hand, a possible hardware structure for the case when only V_B is present is shown in Figure 22.11e, which is similar to a typical hydraulic suspension unit. Further potentially important extensions of the earlier generic configurations may include a consideration of hydraulic inertias, valve/solenoid dynamics, spool resistances, and other related effects, which can be elegantly incorporated with the help of bond graph modeling technique.

22.4 HARDWARE IMPLEMENTATIONS

Despite considerable research into active suspension control designs, only a few systems have emerged. Most of these are prototypes for research and development, while even fewer can be seen in the commercial marketplace. Several of these systems are presented in the following, divided into broad- and narrow-bandwidth categories.

22.4.1 Narrow-Bandwidth Active Suspension

Some narrow-bandwidth systems have been brought to market with limited success. An example of narrow-bandwidth suspension hardware is shown in Figure 22.12a based on [15]. It includes (1) an accumulator (for bandwidth limitation and NVH containment); (2) a mechanical, supporting spring (although this may be absent in some implementations); and (3) a relatively inexpensive pressure-control valve. Structurally, this hardware is similar to the configuration from Figure 22.11e, with the single flow source V_B. Note the single-acting actuator configuration, which is simple and inexpensive.

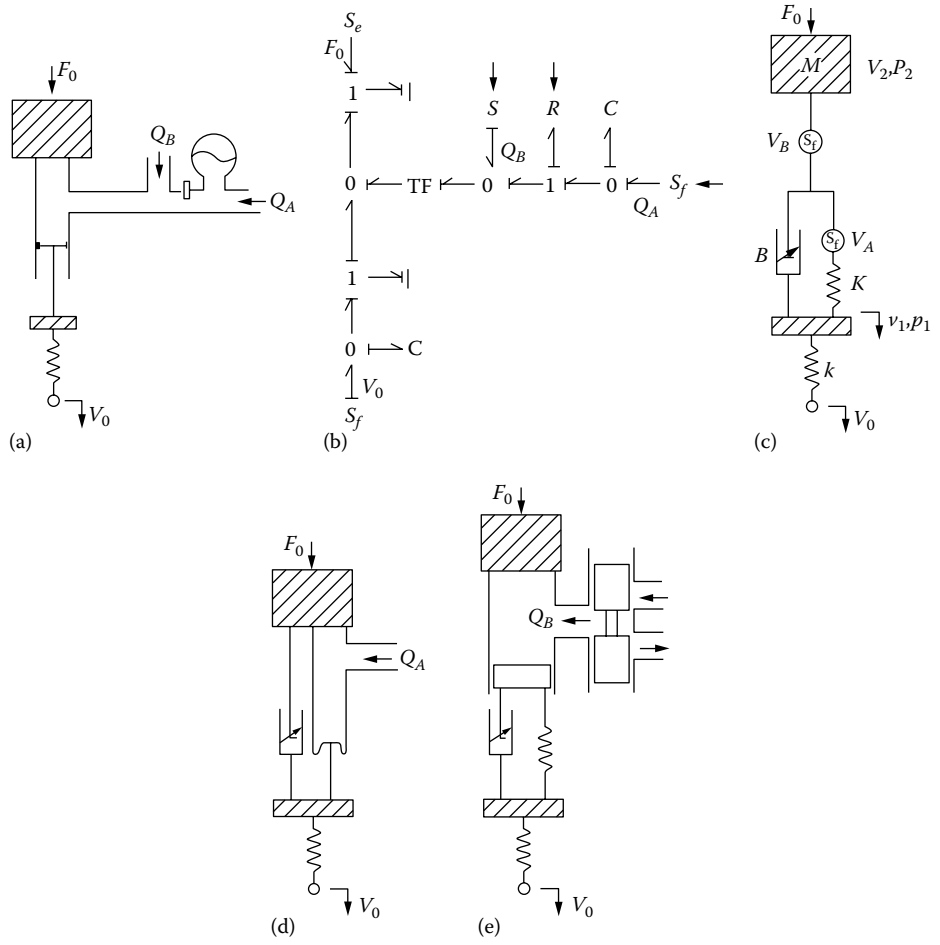


FIGURE 22.11 Narrow-bandwidth suspension configurations and related bond graph. (a) A generic suspension configuration, (b) the equivalent bond graph illustration, (c) the resulting equivalent mechanical structure, (d) example when only V_A is present, and (e) example when only V_B is present.

It does, however, limit the system controllability on the rebound stroke, which is in good part determined by the spring/accumulator settings. While typical, the configuration of Figure 22.12a is not the only one possible and future narrow-bandwidth suspensions may profit from the added flexibility of double-acting actuators and judicious choice of accumulators and parallel springs.

Another hardware alternative would result by replacing the earlier pressure-control valve in Figure 22.12a, with an appropriate flow-control valve. A schematic of a typical proportional, pressure-compensated flow-control valve [16] is shown in Figure 22.12b. It consists of a pressure-control valve augmented by a solenoid-controlled flow metering orifice. Due to a particular interconnection between the two, the valve spool will maintain a constant pressure drop ΔP across the metering orifice, which is preset to a desired value through appropriate solenoid control voltage. This, in turn, will result in a desired output flow, independent of input and output pressure variations.

In the context of narrow-bandwidth suspension actuators, a possible application of the earlier flow-control valve can be rationalized as follows. Any force production with such an actuator will be accompanied with corresponding displacements and velocities of the in-series compliance (cf. Figures 22.11e and 22.12). Thus, in this case, instead of a force control, one can consider the related flow-control problem. However, due to the hardware simplicity, the more direct approach via the pressure-control valves seems to be preferred in practice and is the only one seen in actual production implementations thus far.

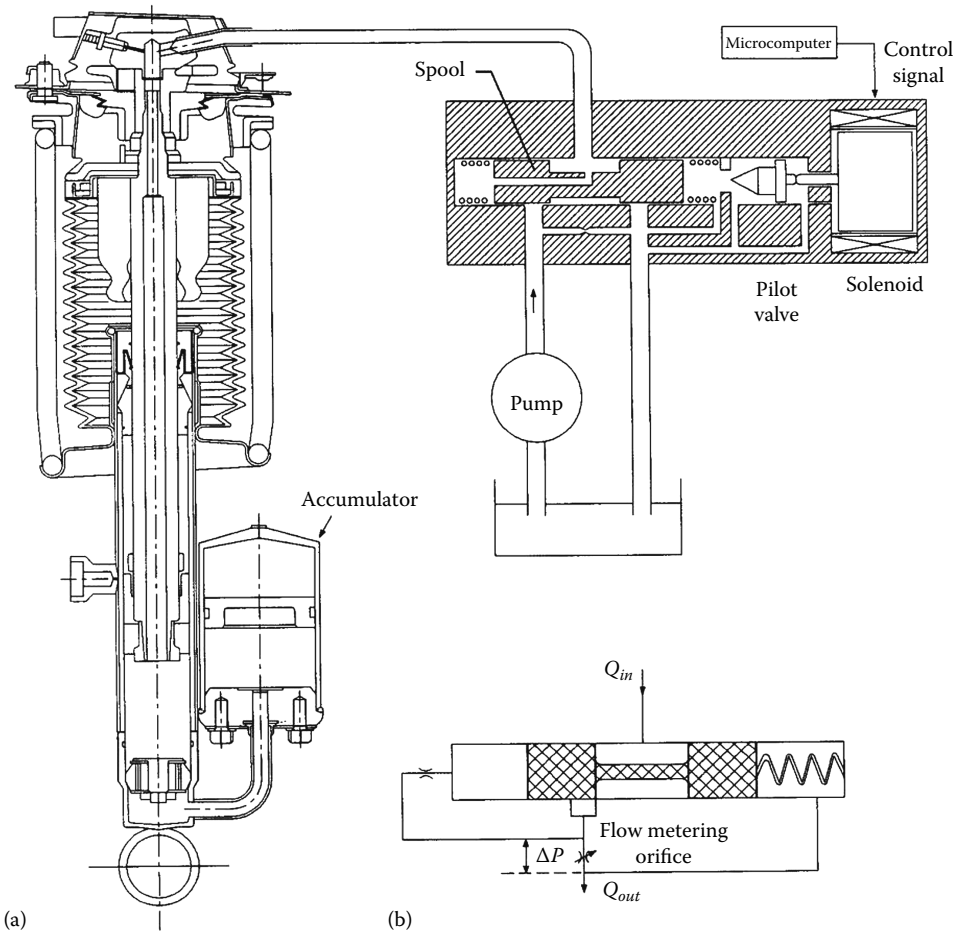


FIGURE 22.12 (a) Schematic diagram of Nissan Infinity Q45, (b) Schematic of a typical proportional, pressure-compensated flow-control valve.

22.4.1.1 Mercedes "Active Body Control"

DaimlerChrysler introduced the "ABC" system on the 2000 CL sedan. The system consists of four low-bandwidth (5 Hz) hydraulic actuators providing fully active load leveling, heave, pitch, and roll control.

Figure 22.13 provides a schematic representation of an ABC suspension system [17,18]. The primary hydraulic cylinder is in series with the steel coil spring. This cylinder sets the spring preload and the series arrangement leads to the low-bandwidth system behavior. In parallel with the spring/cylinder is a twin-pipe gas shock absorber (damper). Hydraulic fluid is supplied to each corner via a pump regulated to 200 bar, and an accumulator at each axle provides pressure leveling during peak loads. High-pressure fluid is carried by two valve blocks—one at each axle. Each block contains two 3-way proportional valves—one for each corner—which are used to maintain plunger pressure and a 2-way shutoff valve for each corner to lock the cylinder when the vehicle is stopped and to provide fail-safe operation. Return fluid is damped through a return accumulator.

The ABC system utilizes multiple sensors for mechanical state measurements. Sensors include longitudinal and lateral accelerometers, three vertical accelerometers (for bumping, pitching, and rolling detections), as well as a level sensor and a plunger position sensor at each wheel. Additionally, several hydraulic pressure sensors are used for pressure control and monitoring.

The ABC system employs various algorithm modules for vehicle dynamic control depending on the vehicle states. Functions provided include speed-based load leveling and ride height adjustment,

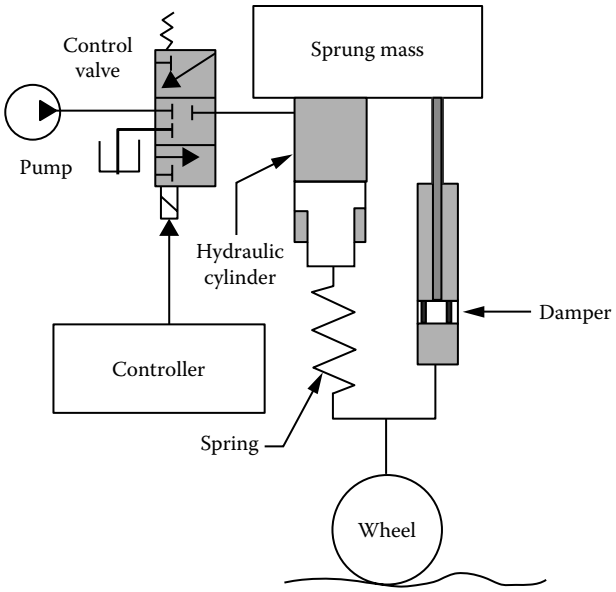


FIGURE 22.13 (a) Schematic diagram of Nissan Infinity Q45, (b) Schematic of a typical proportional pressure-compensated flow-control valve.

skyhook control of the vehicle body heave mode, roll control for roll angle response to vehicle lateral acceleration, and pitch control for pitch angle response to longitudinal acceleration [18]. The algorithms are further controlled by driver input via instrument panel height and performance mode switches.

22.4.2 Broad-Bandwidth Active Suspensions

Since the broad-bandwidth active suspensions (BBASs) attempt to control both the slower body modes and the fast wheel-hop modes, it is in general to be expected that the broad-bandwidth systems will be more difficult to implement. It is expected that such suspensions will incur higher costs and fuel consumption [19]. On the other hand, one could also argue that there is a potential for increased fuel consumption of narrow-bandwidth suspensions due to increased pump flow requirement caused by more compliant hydraulics.

The influence of actuator bandwidth on basic system performance was summarized in [20]. A simplified analysis was done under the assumption that the actuator force production dynamics can be approximated via a simple, first-order, low-pass filter of specified bandwidth. It was shown that the 5 Hz bandwidth actuator can already deliver a performance that is relatively close to the limiting case corresponding to the ideal, infinitely fast actuator. This is in accordance with the analysis in Chapter 31 based on two degree-of-freedom (DOF) quarter-car models, which establishes that most of the ride benefits can be expected around the slow, body mode.

The earlier bandwidth-related study is only an approximate analysis, since it is based on a simple linear model that excludes any coupling influence of vehicle dynamics on actuator force production capabilities. Also, in practice, beyond certain frequencies any suspension performance is dominated by relevant passive components. The actual active suspension force production within the desired bandwidth and road disturbance constraints is a challenging topic in its own right, which is still an open field that may require innovative design solutions as well as full use of mechatronics and associated control system techniques. One possible approach may include the usage of sturdy, polymer-based tug-and-twist actuators proposed by Paynter [21,22].

To establish wide-bandwidth capability of a given actuator–vehicle system in the case of force tracking, one would typically first need to verify that the corresponding force frequency response

is flat up to the cutoff frequency (within 3 dB or 45° phase shift). This should be true for different command force levels and different motions across the actuator mounting points. Once confirmed, the potential benefits of such a full-fledged wide-bandwidth suspension would imply full usage of fast preview that can further improve the overall system performance especially by controlling the wheel-hop mode, as discussed in Chapter 31.

Even if not yet used for production vehicles, this type of suspension can be very useful for R&D work exploring and establishing maximum possible ride, handling, and even active safety benefits of different suspension concepts. It is in this context that most wide-bandwidth suspensions have been (and probably will continue to be) used by different automotive manufacturers and suppliers in their prototype vehicles. An example of a broad-bandwidth suspension that was developed in the late 1980s at Ford [1,23] is shown in Figure 22.1.

The Ford Thunderbird BBAS research vehicle used four high-fidelity electrohydraulic servo actuators, one at each corner of the vehicle. While a typical narrow-bandwidth suspension is controlled through three-way proportional pressure-control valves, the BBAS was controlled through four-way servo valves, which have much higher precision and speed of response. In addition, the BBAS actuators were based on double-acting cylinders being capable of equally fast rebound as well as jounce strokes. The vehicle also had one central and four corner-unit microprocessors for fast signal/control processing, four actuator displacement sensors and four load cells for inner-loop (force) calculations, and four air springs—one at each corner placed in parallel with the BBAS actuators. The air springs support and self-center the vehicle sprung mass as a typical load-leveling set would do and at the same time they provide lower sprung-mass natural frequency for more comfortable basic ride, which is then appropriately dynamically modified through the BBAS actuators. The system incorporated 26 various sensors, including accelerometers, pressure sensors, and vehicle speed sensors.

The BBAS control strategy was based on coordinated individual wheel control and consisted of two hierarchical levels [1,23]. The outer-loop level operated at the 20 ms rate. It calculated the desired corner forces for the four BBAS actuators and desired operating modes (handling or ride dominated) and checked the overall system integrity. The ride-related calculations were based on quarter-car vehicle models aimed at emulating skyhook damping at each corner, which is often very close to optimal possible ride benefit (see Chapter 31). In addition, different spring and damping rates were used depending on prevailing operating modes, that is, ride or handling; Table 22.1 summarizes typical parameters used [23].

The calculations also included roll and pitch attitude control during turning, braking, accelerating, and similar others; this was based on lateral and longitudinal acceleration signals and steering sensors.

TABLE 22.1
Typical Parameters Used in Ford BBAS

		Ride Mode		Maneuvering Mode	
Parameter	**Units**	**Front**	**Rear**	**Front**	**Rear**
Skyhook damping	lb-s/in.	23.5	10.0	23.5	10.0
	N-s/m	4115.3	1751.2	4115.3	1751.2
Conventional damping	lb-s/in.	3.0	3.0	8.0	4.0
	N-s/m	525.4	525.4	1400.9	700.5
Suspension spring	lb/in.	15.0	15.0	120.0	120.0
	N/m	2626.8	2626.8	21,014	21,014

Source: Goran, M.B. and Smith, R.E., Insights gained from active suspension development, *Proceedings of the 26th FISITA Congress*, Prague, Czech Republic, 1996.

The commanded or desired levels of corner forces were communicated to corner computers, which operated at higher sampling rates of 1 ms to implement the inner-loop force control based on feedback provided by load cell sensors. The corner units distributed the force requests between different servo valves. For example, the conventional (relative) suspension damping as well as the passive segments of skyhook force were provided by bypass or cross-port two-way servo valves that provided damping only when called for in a manner similar to SA damping control.

The BBAS R&D vehicle accomplished many of its key objectives, especially with respect to primary ride (up to circa 5 Hz) and handling. The typical results in terms of NASA ride metrics are summarized in [23], where a smaller number on the scale of 10 implies better ride. From this reference it can be seen that, on average, the active Thunderbird provided the best ride among a dozen different vehicles under consideration. The only areas where the BBAS had somewhat worse performance were on smooth roads, where Jaguar XJ6 and Lexus LS 400 (both well known for their smooth ride) had an advantage. This may be attributable to still less refined NVH (higher-frequency) levels that are a remaining challenge for BBAS concepts. Based on results summarized in [23], it should also be noted that the BBAS Thunderbird significantly outperformed both the baseline passive suspension Thunderbird and its narrow-bandwidth (NBAS) counterparts as exemplified by production Infinity Q45a and Toyota Soarer.

The handling characteristics of the active Thunderbird are compared with its passive (production) counterpart in [23]. The results of back-to-back tests show that the active vehicle results in more agile handling, with faster yaw response, less roll and pitch, and significantly larger maximum lateral acceleration (0.895 vs. 0.81 g). Thus, in a direct comparison between the two vehicles, the active car displayed significant improvements in both vehicle ride and handling.

While achieving the BBAS and all their potential benefits in the strict sense of broad-bandwidth suspensions still remains an illusive goal, a significant progress has been achieved with the previous early efforts. This holds particularly for the case of primary ride and handling. The challenging areas in need of further development include secondary ride, NVH, and system power consumption, complexity, and cost. Possible approaches here include introduction of tuned vibration absorbers for wheel-hop mode and addition of controlled in-series compliance, which if properly designed would provide appropriate NVH isolation. This also further reduces the distinction between the BBAS and their NBAS counterparts.

22.4.3 Semiactive Suspensions

One of the first production SA suspensions was introduced in the 1987 Thunderbird Turbo Coupe employing electronically controlled damping [24]. The damping is adjusted according to driving conditions such as during hard cornering, braking, acceleration, and at high speeds. The damping adjustment is accomplished by opening a rotary shutter valve to allow oil to bypass the piston valve (see Figure 22.14). Check valves in the shutter assembly further allow the damping at the same setting/rotating position to be different for compression and extension.

22.5 VEHICLE MODELS

In this section, we briefly review different vehicle models that are used for the design of optimal suspension ride control algorithms. We start with simple, quarter-car models, which allow for only 1D, vertical, or heave motion. Within this setting, the simplest possible one DOF model consists of just the vehicle or sprung mass, which, as shown in Figure 22.15, is supported by a suspension actuator positioned between the vehicle and ground. Moreover, it is a common practice to assume that the vehicle is moving forward with constant velocity V, so that the road roughness disturbance appears as a vertical input proportional to V and the spatial slope of the road unevenness. It is in this context that we refer to the so-called "moving" ground.

Adding a wheel/tire subassembly (unsprung mass) to the earlier one DOF structure leads to two DOF, quarter-car models as shown, for example, in Figure 22.16. The unsprung-mass mode is often referred to as "wheel hop" and is characterized by relatively light damping and a natural frequency

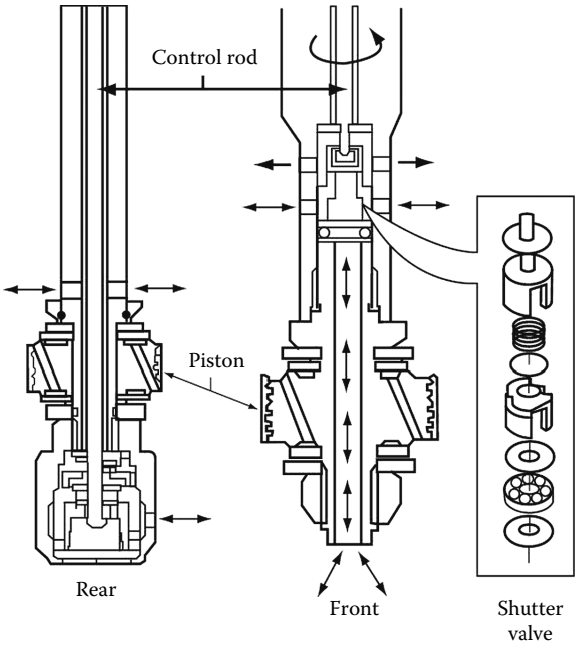


FIGURE 22.14 Adjustable shock absorber used in 1987 Thunderbird Turbo Coupe. (From Soltis, M.W., Thunderbird turbo coupe programmed ride control (PRC) suspension, SAE Paper 870540, 1987.)

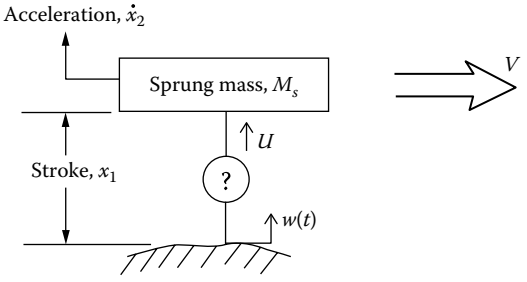


FIGURE 22.15 A one DOF model.

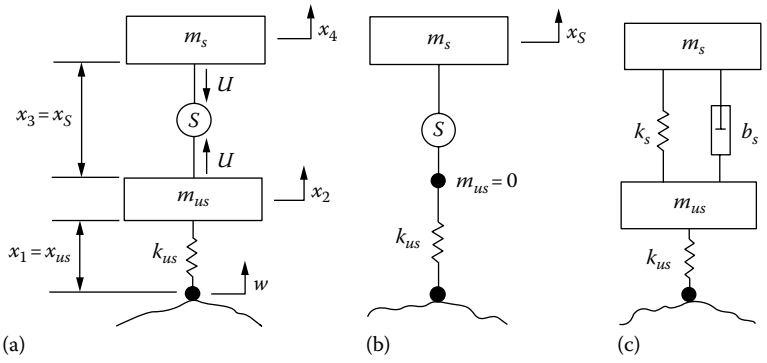


FIGURE 22.16 A two DOF model: (a) with active suspension, (b) corresponding limiting case, (c) with conventional passive suspensions.

between 8 and 12 Hz. The principal body mode, or sprung-mass mode, is typically around 1 Hz, with active suspensions leaning toward lower, 0.5 Hz range values. The related state–space equations are summarized in Appendix 22.A. A justification for the simple two DOF model can be found in [25]. The authors use detailed ADAMS™-based vehicle models and step responses to identify the two DOF model parameters. For the typical McPherson strut suspension shown in Figure 22.17, it was found that using the simple component-based data (e.g., the unsprung mass would typically consist of corresponding wheel, knuckle, and attached linkages) would suffice. However, for some other suspension configurations—such as double-wishbone suspensions—there could be significant differences between the ADAMS™ identified data and the component-based data due to differences in corresponding kinematic paths during typical ground input events.

Typical half-car, 2D models include sprung-mass pitch and heave or vertical modes (Figure 22.18), often augmented by front and rear unsprung-mass vertical dynamics and possibly engine/transmission block dynamics. The full-car, 3D models, on the other hand, include vehicle roll as well as the pitch and heave modes (Figure 22.19). These models consider only rigid body motions and are based on lumped-parameter, linearized dynamics, which are well suited for applications of the LQG optimal control technique.

A typical automotive ride dynamic model is most often confined to close-to-linear operation characterized by relatively small sprung-mass velocities and displacement angles and suspension and tire deflections. To what extent this assumption remains satisfied is best evaluated through actual vehicle tests. Some useful related results can be found in [26]. The authors compare

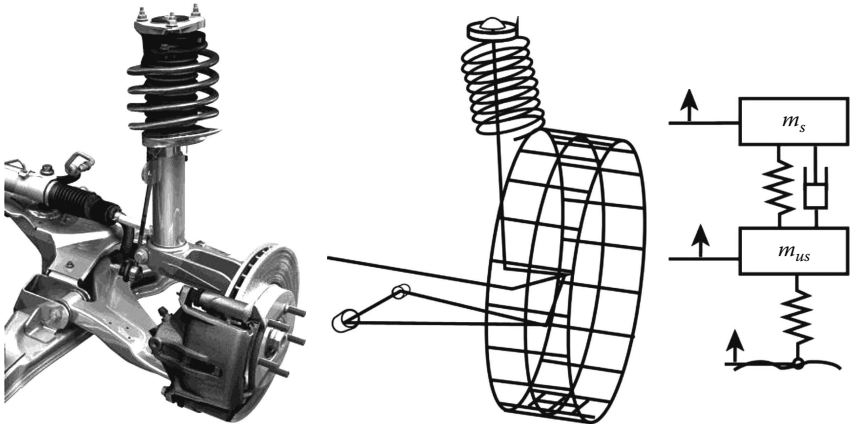


FIGURE 22.17 McPherson strut suspension and its equivalent two DOF model.

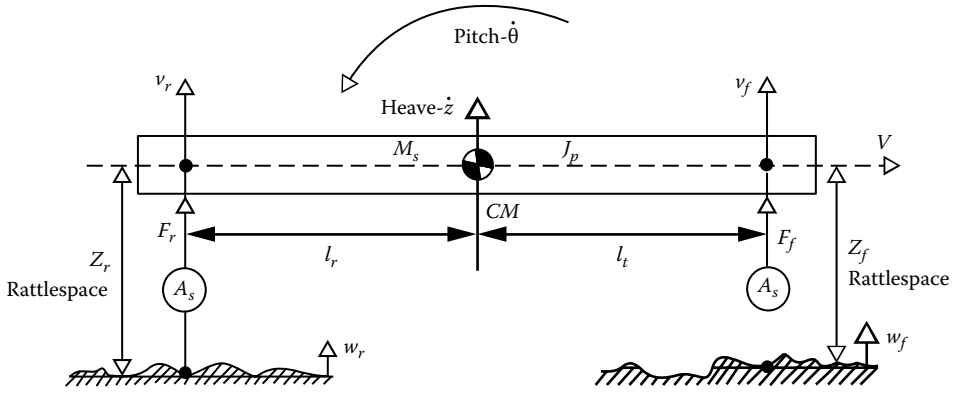


FIGURE 22.18 Half-car two DOF model.

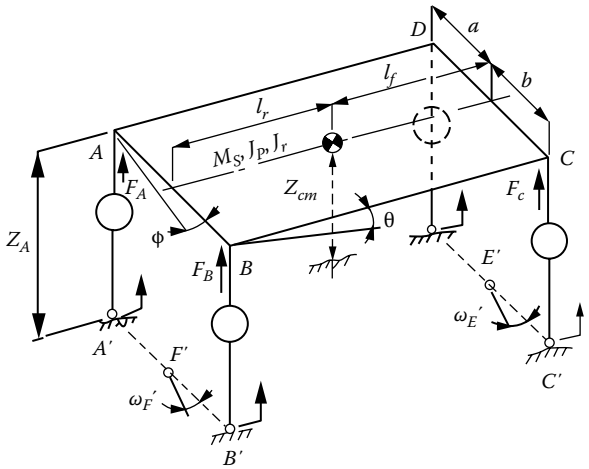


FIGURE 22.19 3D vehicle model schematic.

measured and predicted vertical accelerations at the driver's seat location for the same measured road input sequence. The prediction is based on a linear, seven DOF vehicle model of a 1974 Buick Luxus with conventional, passive suspension. The seven DOFs were made to correspond to rigid body heave, pitch, and roll, and four DOFs for unsprung-mass dynamics (the vertical displacements of each of the four wheels). The results of this study [26] showed a close match up to frequencies of about 10 Hz, with discrepancies above 10 Hz being primarily due to unmodeled high-frequency elastic modes and other neglected phenomena, such as tire unbalance and engine-block dynamics.

The same reference also compared linear vehicle models of different complexity. This was done through a comparison of predicted acceleration responses for the quarter-car, two DOF; half-car, four DOF; and the earlier full-car, seven DOF models. It was concluded that the difference between the various models is relatively small, especially when the difference between predicted and measured traces is considered. This gives credence to the usage of simpler, linear models, particularly when used in preliminary studies that, as in the present case, serve to establish the prevailing trends and benefits.

Another set of comparisons with measured data was pursued in [27], where a linear 2D, single-track, six DOF vehicle model was used, consisting of sprung-mass heave and pitch, front and rear unsprung-mass bounce, and engine and passenger dynamics. The results in terms of measured and predicted center of gravity C.G. acceleration power spectrum densities (p.s.d.'s) and left front spindle (unsprung-mass) acceleration p.s.d.'s are reproduced in Figures 22.20 and 22.21, respectively. The data were obtained with a 1979 Ford Fairmont traversing Ford's Dearborn Proving Ground straightaway at 55 mph. For the analysis, the measured road spectrum was approximated via appropriate exponential/log functions. Comparing the measured and predicted acceleration, it can be concluded that the agreement is in general very good for frequencies up to about 12 Hz, with the discrepancy at higher frequencies being again attributed to unmodeled, higher-frequency flexible modes, tire unbalance, and various nonlinearities.

The linear, low-order models are in general adequate for system-level studies such as optimal suspension control synthesis and analysis that will be discussed in Chapter 31. Such system-level studies serve as a precursor necessary to establish the potential benefits and overall trends and R&D directions. Once a candidate suspension has been established in this way, further evaluations would typically proceed via more refined and complex models that would include flexible modes, actuator dynamics, and various nonlinear effects. Since ride and handling are the attributes with many inherent subjective components, final evaluation and verdict still implies the need for actual vehicle test drives and verification.

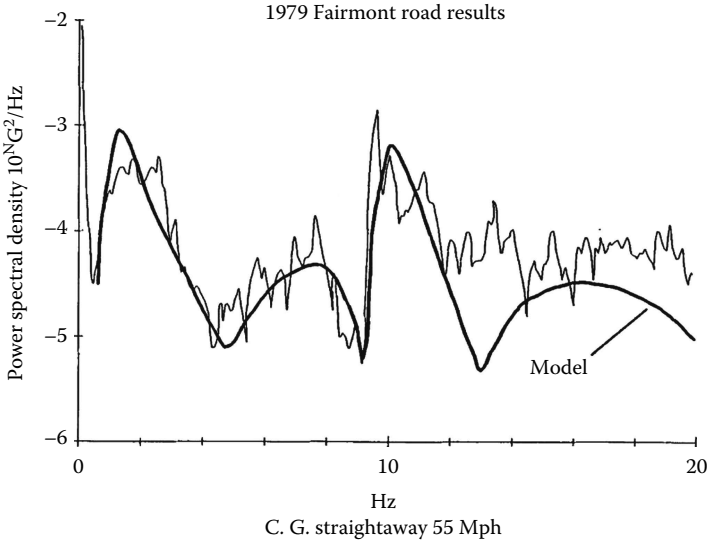


FIGURE 22.20 Comparison between measured and predicted CG acceleration p.s.d.'s for a linear six DOF model. (From Smith, R.E. and Sigman, D.R., Experimental verification of a linear rigid body model, Ford Motor Company Research Report, Dearborn, MI, October 1981.)

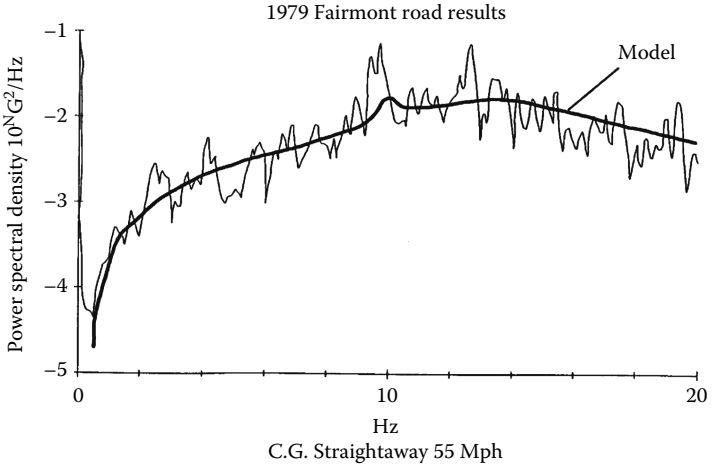


FIGURE 22.21 Comparison between measured and predicted unsprung-mass acceleration p.s.d.'s for a linear six DOF model. (From Smith, R.E. and Sigman, D.R., Experimental verification of a linear rigid body model, Ford Motor Company Research Report, Dearborn, MI, October 1981.)

22.6 OTHER CONCEPTS

22.6.1 Anti-Dive, Anti-Squat, and Anti-Pitch Suspension Designs

In passive suspensions, geometry may be used to modify the load transmission paths so that the weight transfer associated with longitudinal acceleration and braking is only partially supported through suspension travels. Under proper design measures, the complementary amount of load then results in reduced suspension travel or "anti-suspension travel"; this amount is supported through suspension linkages instead of suspension spring deflection and travel. The amount of anti-suspension travel consequently effects the vehicle attitude and is referred to as anti-squat (usually associated with acceleration), anti-dive (usually associated with braking), or anti-pitch, which is often used as a general term for vehicle attitude correction.

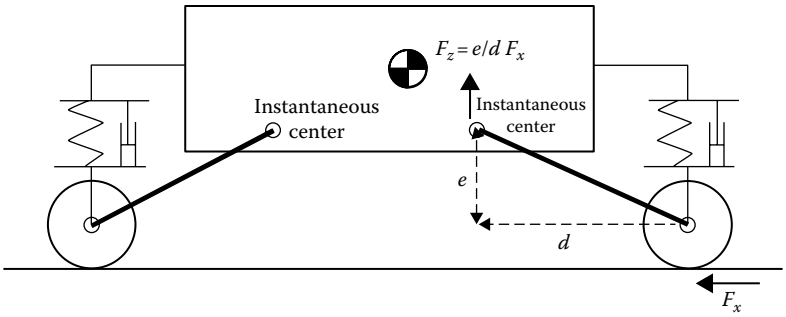


FIGURE 22.22 Illustration of equivalent trailing arm and anti-squat geometry for a rear-wheel-drive, independent rear suspension vehicle.

The mechanics of the anti-x features can be explained through an "equivalent trailing arm analysis" (see Figure 22.22) that recognizes that all suspensions are functionally equivalent to a trailing arm [28]. The mechanics can also be explained through an "equivalent compliance analysis" [29,30] that translates the suspension geometry, kinematics, and bushing compliance into an equivalent compliance regarding the reaction forces and moments onto the vehicle.

Active suspensions, on the other hand, can insulate the vehicle from both road irregularities and load disturbances (e.g., inertial load induced by braking, accelerating, or cornering), independently of each other, something that is not possible with their passive counterparts. An active suspension design that improves the load responses while maintaining the ride comfort characteristic of a soft passive suspension is discussed in [31]. An active suspension control designed with certain amount of anti-x features in mind can further be combined with the more common approach of anticipation or feedforward control to counteract gas/brake pedal applications. In any case, the amount of existing passive and/or active anti-x and their effects should be taken into account when calculating feedforward control.

22.6.2 Delft Suspension

The so-called Delft suspension represents an original design approach toward practical active or actively modulated suspension [32–34]. Conceptually, the overall suspension force is controlled via the mechanism shown in Figure 22.23. The suspension mechanism force F_C is controlled by changing the adjustable lever distance "*b*" in Figure 22.23. This concept was realized through a 3D contraption comprising of a cone mechanism with one end of the spring *S* from Figure 22.23 attached to the vehicle sprung mass at point A and the other to an adjustable crank mechanism regulated by an electrical servomechanism. The crank rotation causes the spring to rotate along the surface of an imaginary cone, with the length of the spring remaining theoretically constant. The net effect is that the force F_C from Figure 22.23 is now continuously controlled through the servo–crank mechanism rotations, with ideally negligible power consumption.

To date, good success was reported with the attitude control during cornering and braking. However, in practical implementation, the power consumption was still significant [20], and limited ride tests reported thus far show about 10% improvements in the rms accelerations, which are primarily concentrated around the dominant, sprung-mass oscillation mode. Since the earlier concept embeds the characteristics of a variable (SA) spring suspension, it is expected to have significant practical ride improvement potentials, which may be materialized through further developments in the future.

22.6.3 Roll-Control Suspensions

While fully active suspensions imply certain cost and complexity issues, as a compromise solution to improved ride and handling with moderate cost and complexity increases, a number of automotive manufacturers have been introducing active and SA suspensions that are focused on vehicle roll control only. Here a significant decrease in the power required can be achieved—along with performance

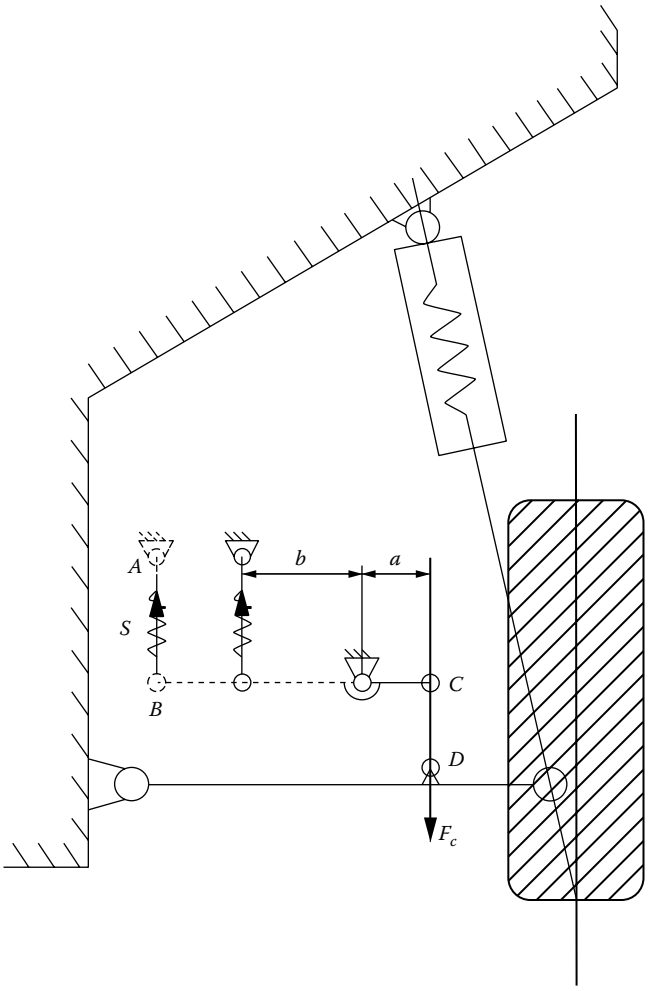


FIGURE 22.23 Delft suspension mechanism with variable lever "*b*." (Based on Venhovens, P.S.Th. et al., Semiactive vibration and attitude control, *Proceeding of International Symposium on Advanced Vehicle Control*, Yokohama, Japan, pp. 170–175, 1992.)

improvement—by feeding forward driver's action as reflected through steering wheel, brake, and accelerator pedal inputs. It is well known that a good ride comfort can be achieved with softly tuned passive or SA suspension, with corresponding soft roll bar settings. However, this will result in large pitch and roll motions during more aggressive braking and cornering, respectively. The active (or SA) roll control is then used to satisfy these conflicting constraints. In practice, the concept is typically realized by inserting a small hydraulic jack and a gas spring accumulator with a control solenoid, all placed in series with the anti-roll bar. The anti-roll bar can be as stiff as in street race cars, without negatively impacting the straight-line ride since the solenoid connecting the hydraulic jack and the accumulator is then kept open, thus lowering the effective stiffness of the roll bar—accumulator combination. This situation is shown in the upper portion of Figure 22.24.

During turns, however, the solenoid is closed, so that the full stiffness of the beefed-up anti-roll bar acts to prevent excessive vehicle roll—see the middle portion of Figure 22.24. In the case of extreme turns, the anti-roll action can be further amplified by directing high-pressure oil into the appropriate hydraulic jack cylinder as shown in the lower portion of Figure 22.24. A production implementation of this concept can be found in 1995 Citroen Xantia Activa [35], where it effectively augments the already established ride capabilities of their Hydractive or hydropneumatic load-leveling suspension [36].

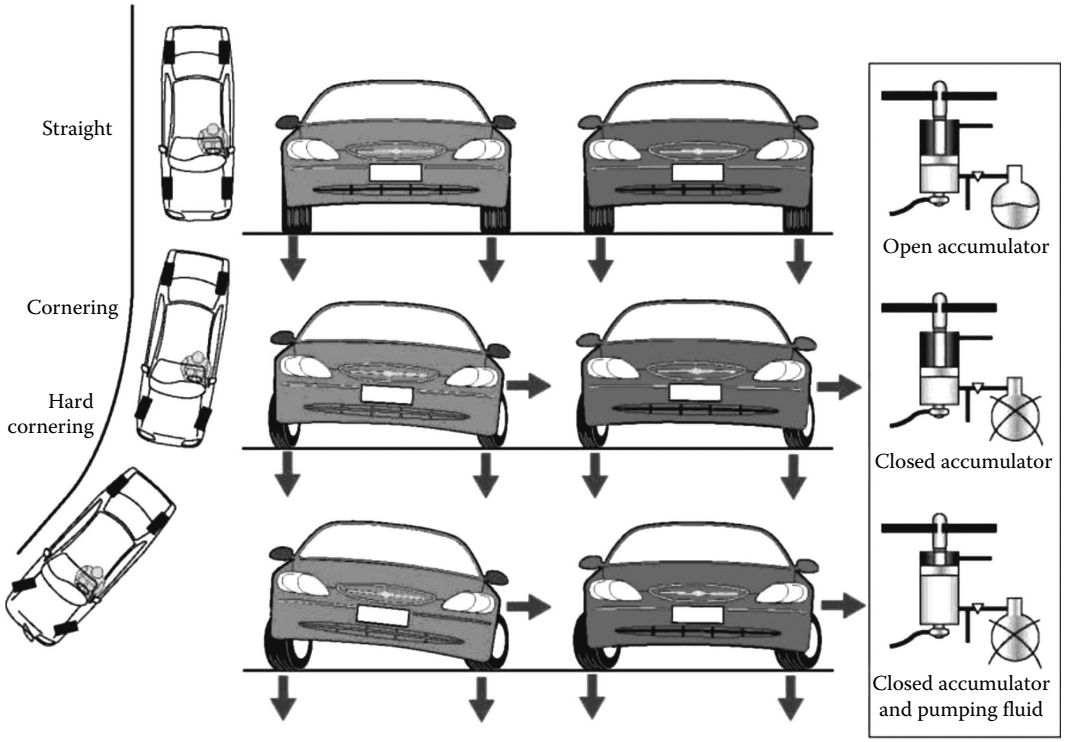


FIGURE 22.24 Roll control during straight-line driving and cornering.

The benefits of roll control were also discussed in [37], where increased anti-roll moments have been achieved with the help of a hydraulic rotary actuator inserted around the middle of the stabilizer bar. A similar type of rotary actuators was fitted at the front and rear axles of some recent higher-end BMW vehicles, which led to the "Dynamic Drive" concept described in [38]. The front and rear roll-control actuators are controlled by means of two electronically regulated pressure-control valves, which minimize the roll motion in corners and road input roll disturbances, while at the same time they contribute to overall vehicle steering accuracy and agility. Future benefits of active roll control will be focused around increased functionality and improvements in overall vehicle stability and active safety through coordinated action with other vehicle control systems, such as brakes, active steering, and electronically controlled all-wheel drive. This can be particularly useful in emergency maneuvers such as sudden need for vehicle stabilization and control during evasive actions on slippery road surfaces.

22.6.4 Dynamic Vibration Absorbers or Tuned Mass Dampers

In its most elementary form, DAs or tuned mass dampers consist of a mass–spring–damper oscillator system tuned to "absorb" the excessive oscillations at a target natural frequency. Typically, this target resonance is characterized by lack of adequate damping, so that in effect the DA here serves to augment the critical resonance mode damping. To date the only known production application of DA in the context of suspensions could be found on the Citroen 2CV—a small car that has acquired a "cult" status in view of its many original characteristics including the superior ride.

While the DA mass constitutes only a small fraction of the target resonance mass, it still adds weight, so in order to justify this additional device, it is important to clearly establish the related benefits. This was done in [39]. From there it can be seen that the DA can nicely complement the active suspension and facilitate full realization of its potentials. This is further enhanced with active DA means for further substantial increase in potential benefits [20,39].

22.6.5 Additional Considerations

While the main emphasis in this chapter was on secondary or sprung-mass suspension, it is also possible to introduce advanced active or SA suspension elements at other vehicle locations. Examples include isolation of vehicle seats and heavy-duty truck cabins [20]. These can further enhance the protection of vehicle occupants and cargo from road-/drive-induced vibrations, especially if combined with the previously mentioned sprung- and unsprung-mass suspensions.

Finally, it should be mentioned that most if not all of production active suspension units introduced to date are of narrow-bandwidth variety (Section 22.4.1). However, future progress in hardware and microelectronics may facilitate full exploitation of broad-bandwidth suspension capabilities. These include exploiting the potentials of preview controls discussed in Chapter 31, as well as implementing rapid and proactive interventions in case of sudden, safety-related maneuvers. It is conceivable that future advanced suspensions could include fast unsprung actuators in tandem with adaptive sprung-mass actuators with gradually varying bandwidth; this could be achieved through interchangeable sets of SA springs and dampers. This is an example of a judicious suspension structure based on sound control/systems insight and principles. More about the latter can be found in Chapter 31.

22.A APPENDIX

A two DOF quarter-car model equipped with a hydraulic active suspension actuator is presented here. With the assumptions that the springs and damping are linear and the tire damping is negligible and with the help of bond graph analysis (see Figure 22A.1), the active suspension system can be represented mathematically as

$$\dot{x} = Ax + Bf + B_w\dot{w}$$

where the four states x_1–x_4 are defined as in Figure 22.16a,

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\omega_{us}^2 & -2\zeta\rho\omega_s & \rho\omega_s^2 & 2\zeta\rho\omega_s \\ 0 & -1 & 0 & 1 \\ 0 & 2\zeta\omega_s & -\omega_s^2 & -2\zeta\omega_s \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ \rho \\ 0 \\ -1 \end{bmatrix}, \quad B_w = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\omega_s = \sqrt{\frac{k_s}{m_s}}$$

$$\omega_{us} = \sqrt{\frac{k_{us}}{m_{us}}}$$

$$\zeta = \frac{b_s}{2m_s\omega_s}$$

$$\rho = \frac{m_s}{m_{us}}$$

where
m_s is the sprung mass
m_{us} is the unsprung mass
k_s is the suspension spring constant
k_{us} is the tire spring constant
b_s is the suspension damping coefficient
z_s is the sprung-mass displacement
z_{us} is the unsprung-mass displacement
f is the active actuator force
w is the road disturbance displacement, as shown in Figure 22A.1

The hydraulic actuation system (see Figure 22A.2) can be described, based on servo-valve dynamics [16,40], as

$$\dot{f} = \frac{\sqrt{2A_p}\beta k_{xd}}{V} x_{sp} \cdot \text{signsqrt}\left(\frac{P_s - \text{sign}(x_{sp})f}{A_p}\right) + \frac{2A_p^2\beta}{V}(\dot{z}_{us} - \dot{z}_s)$$

$$\dot{x}_{sp} = \frac{1}{\tau}(-x_{sp} + k_{sv} i_{sv})$$

where
A_p is the piston area
β is the fluid bulk modulus
k_{xd} is the orifice flow coefficient
x_{sp} is the servo-valve displacement
$\text{signsqrt}(y) \equiv \text{sign}(y)\sqrt{|y|}$
P_s is the supply pressure
V is the cylinder chamber volume
τ is the time constant for servo-valve displacement
k_{sv} is the valve gain

If we linearize the hydraulic equation around the servo-valve operating point, we have

$$2C_x x_{sv} + 2A_p\left(\dot{z}_{us} - \dot{z}_s\right) = \frac{V}{\beta A_p}\dot{f}$$

$$\tau\dot{x}_{sv} + x_{sv} = K_v i_{sv}$$

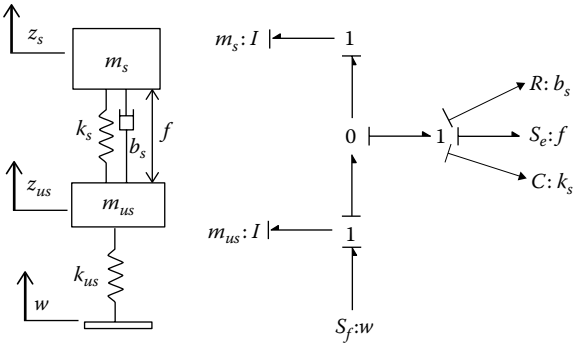


FIGURE 22A.1 Two DOF quarter-car model and its bond graph.

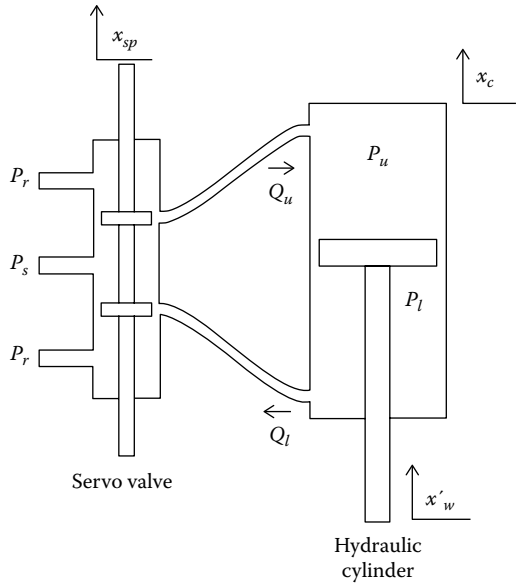


FIGURE 22A.2 Hydraulic servo valve.

And the linear quarter-car active suspension model can be extended [40] as

$$\frac{d}{dt}\begin{bmatrix} z_{us}-w \\ \dot{z}_{us} \\ z_s-z_{us} \\ \dot{z}_s \\ f \\ x_{sv} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\omega_{us}^2 & -2\zeta\rho\omega_s & \rho\omega_s^2 & 2\zeta\rho\omega_s & \dfrac{-\rho}{m_s} & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 2\zeta\omega_s & -\omega_s^2 & -2\zeta\omega_s & \dfrac{1}{m_s} & 0 \\ 0 & \dfrac{2\beta A_p{}^2}{V} & 0 & -\dfrac{2\beta A_p^2}{V} & 0 & \dfrac{2\beta A_p^2}{V}C_x \\ 0 & 0 & 0 & 0 & 0 & \dfrac{-1}{\tau} \end{bmatrix} \begin{bmatrix} z_{us}-w \\ \dot{z}_{us} \\ z_s-z_{us} \\ \dot{z}_s \\ f \\ x_{sv} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \dfrac{k_v}{\tau} \end{bmatrix} i_{sv} + \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \dot{w}$$

REFERENCES

1. Goran, M.B., Bachrach, B.I., and Smith, R.E.: The design and development of a broad bandwidth active suspension concept car. SAE Paper No 925100, *Proceedings of the 24th FISITA Congress,* London, U.K., 1992, pp. 231–252.
2. Davis, R.I. and Patil, P.B.: Electrically powered active suspension system for a vehicle. U.S. Patent: 5,060,959 (1992).

3. Miller, J.M. and Grabowski, J.R.: Current mode hysteresis controller for pulse width modulated inverters. U.S. Patent 5,159,542 (1992).
4. Hrovat, D., Asgari, J., and Fodor, M.: Automotive mechatronic systems. In: Leondes, C.T.: *The Mechatronic Systems Techniques and Applications*, Gordon and Breach Science Publishers, Amsterdam, the Netherlands, 2000, pp. 1–98.
5. Moran, T.: A new suspension's magnetic appeal. *The New York Times*, October 11, 2004, p. D10.
6. Bodie, M. and Hac, A.: Closed loop yaw control of vehicles using magneto-Rheological Dampers, *SAE Technical Papers*, 2000-01-0107, 2000, doi:10, 4271/2000-01-0107.
7. Fodor, M. and Redfield, R.C.: Experimental verification of resistance control, semi-active damping, *Vehicle System Dynamics*, 26(2) (1996): 143–159.
8. Chance, B.K.: Continental mark VII/Lincoln continental electronically-controlled air suspension system, SAE 840342, 1984.
9. Watton, J., Holford, K.M., and Surawattanawan, P.: Electrohydraulic effects on the modelling of a vehicle active suspension, *Proceedings of the Institution of Mechanical Engineers*, *Part D: Journal of Automobile Engineering*, 215 (2001): 1077–1092.
10. Hrovat, D. and Hubbard, M.: A comparison between jerk optimal and acceleration optimal vibration isolation, *Journal of Sound and Vibration*, 112(2) (1987): 201–210.
11. Alleyne, A. and Liu, R.: On the limitations of force tracking control of hydraulic servosystems, *ASME Journal of Dynamic Systems, Measurement, and Control*, 121(2) (1999): 184–190.
12. Zhang, Y. and Alleyne, A.: A practical and effective approach to active suspension control, *Proceeding of the 6th International Symposium on Advanced Vehicle Control*, Hiroshima, Japan, September 2002, pp. 153–158.
13. Peng, H., Strathearn, R., and Ulsoy, A.G.: A novel active suspension design technique—Simulation and experimental results, *Proceedings of the 1997 American Control Conference*, Vol. 1, Albuquerque, NM, 1997, pp. 709–713.
14. Karnopp, D.C.: Active suspensions based on fast load levelers, *Vehicle System Dynamics*, 16 (1987): 355–380.
15. Akatsu, Y., Fukushima, N., Talcahashj, K., Satch, M., and Kawaranki, Y.: An active suspension employing an electrohydraulic pressure control system, SAE Paper No. 905123, 1990.
16. Merritt, H.E.: *Hydraulic Control Systems*. Wiley, New York, 1967.
17. Merker, T., Gaston, G., and Olaf, T.: Active body control (ABC) the DaimlerChrysler Active Suspension and Damping System, *SAE Technical Paper* 2002-21-0054.
18. Merker, T., Wirtz, J., Hill, M., and Jeglitzka, M.: Das SL Fahrwerk-Dynamik und Komfort vereint (The SL Chassis-Dynamics and Comfort Combined). *ATZ/MTZ Magazine Special Issue* (2001), 84–91.
19. Hillebrecht, P., Konik, D., Pfeil, D., Wallenowitz, H., and Zieglmeier, F.: The active suspension between customer benefit and technological competition. SAE Paper No. 925099, *Proceedings of the 24th FISITA Congress*, London, U.K., 1992, pp. 221–230.
20. Hrovat, D.: Survey of advanced suspension developments and related optimal control applications, *Automatica*, 33(10) (1997): 1781–1817.
21. Paynter, H.M.: High-pressure fluid-driven tension actuators and method for making them. U.S. Patent 4,751,869 (1988).
22. Paynter, H.M.: Thermodynamic Treatment of tug-&-twist technology, *Proceedings of the 1996 Japan–U.S.A. Symposium on Flexible Technology*, Boston, MA, 1996.
23. Goran, M.B. and Smith, R.E.: Insights gained from active suspension development, *Proceedings of the 26th FISITA Congress*, Prague, Czech Republic, 1996.
24. Soltis, M.W.: Thunderbird turbo coupe programmed ride control (PRC) suspension, SAE Paper 870540, 1987.
25. Kim, C., Ro, P.I., and Kim, H.: Effect of the suspension structure on equivalent suspension parameters, *Proceedings of the Institution of Mechanical Engineers*, 213(Part D) (1999): 457–470.
26. Healey, A.J., Nathman, E., and Smith, C.C.: An analytical and experimental study of automobile dynamics with random roadway inputs, *ASME Journal of Dynamic Systems, Measurement, and Control*, 99(4) (1977): 284–292.
27. Smith, R.E. and Sigman, D.R.: Experimental verification of a linear rigid body model, Ford Motor Company Research Report, Dearborn, MI, October 1981.
28. Gillespie, T.D.: *Fundamental of Vehicle Dynamics*. ISBN 1-56091-119-9, SAE, Warrendale, PA, 1992.
29. Tseng, H.E., Kuang, M., Asgari, J., and Hrovat, D.: Chassis modeling of passenger vehicle for clutch engagement/gear shift events, *Proceeding of the 4th International Symposium on Advanced Vehicle Control*, Nagoya, Japan, September 1998, pp. 123–128.

30. Tseng, H.E., Asgari, J., and Hrovat, D.: Modeling and analysis for simulation of vehicle chassis motion during powertrain engagement/gear shift events, *Proceeding of 1998 Western Multi-Conference, International Conference on Simulation and Multimedia in Engineering Education*, San Diego, CA, January 1998, pp. 280–290.
31. Wang, F. and Smith, M.C.: Active and passive suspension control for vehicle dive and squat. In: Johansson, R. and Rantzer, A. (eds.): *Nonlinear and Hybrid Systems in Automotive Control*, ISBN 1-85233-652-8, Springer-Verlag, London, U.K., 2003.
32. van der Knapp, A.C.M.: Design of a low power anti-roll/pitch system for a passenger car. Delft University of Technology, Vehicle Research Laboratory, Delft, the Netherlands, Report 89.3VT.2628.
33. van der Knapp, A.C.M. and Paceijka, H.B.: Mass spring system with roll/pitch stabilization for use in vehicles. International Patent Application PCT/NL93/00094.
34. Venhovens, P.S.Th., van der Knaap, A.C.M., and Pacejka, H.B.: Semiactive vibration and attitude control, *Proceeding of International Symposium on Advanced Vehicle Control*, Yokohama, Japan, 1992, pp. 170–175.
35. Anonymous: Roll-Free Cornering, *Popular Science*, 247(1) (1995): 28.
36. Carbonaro, O.: Hydractive suspension electronic control system, control technology and philosophy, *Proceedings of the 23rd FISITA Conference*, Paper 905101, Torino, Italy, 1990.
37. Reusing, O., Ochs, M., Walz, U.G., and Bungeler, J.: Low-cost active suspension system, *Proceedings of International Symposium on Advanced Vehicle Control*, Yokohama, Japan, 1992, pp. 176–180.
38. Konik, D., Bartz, R., Barnthol, F., Bruns, H., and Wimmer, M.: Dynamic drive—The new active roll stabilization system from the BMW Group—System description and functional improvements, *Proceeding of International Symposium on Advanced Vehicle Control,* Ann Arbor, MI, 2000, pp. 413–420.
39. Hrovat, D.: Optimal active suspension structures for quarter-car vehicle models, *Automatica*, 26(5) (1990): 845–860.
40. Shen, X. and Peng, H.: Analysis of active suspension systems with hydraulic actuators, *Vehicle System Dynamics*, 41 (2004): 143–152.

23 Driveline

Andreas Laschet and Ferit Küçükay

CONTENTS

23.1 Introduction 798
23.2 Procedure of the Analysis of Dynamic Effects in a Driveline 798
23.3 Torsional Vibration Models of a Driveline 798
23.4 Analysis of Excitability 802
23.4.1 Basics 802
23.4.2 Sensitivity of the Natural Behavior 803
23.4.3 Resonance Diagram 804
23.5 Components of a Driveline 806
23.5.1 Engine 807
23.5.2 Flywheel and Clutch 808
23.5.2.1 Torsional Vibration Damping at the Transmission Input 808
23.5.2.2 Starting Devices 811
23.5.2.3 Mechanical Clutches 811
23.5.2.4 Hydrodynamic Torque Converter 812
23.5.3 Manual and Automatic Transmission 814
23.5.3.1 Manual Transmission 816
23.5.3.2 Automated Manual Transmission 816
23.5.3.3 Continuously Variable Transmission 817
23.5.3.4 Double-Clutch Transmission 817
23.5.3.5 Automatic Transmission 818
23.5.4 Drive Shafts 820
23.5.4.1 Rear-Wheel Drive and Front-Wheel Drive 821
23.5.4.2 Four-Wheel Drive and All-Wheel Drive 822
23.5.4.3 Cardan (Universal) Shaft 822
23.5.5 Differential Gear 822
23.6 Simulation of the Dynamic Driveline Behavior: Examples 823
23.6.1 Drivelines: Studying Major Dynamic Effects in a Vehicle 823
23.6.2 Nonlinear Behavior of Drivelines 824
23.6.2.1 Introducing a Simplified Calculation to Evaluate Gear Rattling 824
23.6.2.2 Simulation Models to Analyze the Gear Rattle Behavior 828
23.6.2.3 Advanced Evaluation of Superimposed Nonlinear Characteristics 829
23.6.3 Driveline Modeling and Simulation of the Shifting Behavior 829
23.6.3.1 Targets and Benefits of Shifting Simulation 829
23.6.3.2 Development of Shifting Simulation Methods 831
23.6.3.3 Example: Manually Shifted Powertrain 832
23.6.3.4 Example: Dual-Clutch Transmission 832
23.7 Summary 833
Acknowledgments 834
References 834

23.1 INTRODUCTION

Since Aristotle (Greek philosopher 384–322 BC), the statement has been valid that the "whole is greater than the sum of its parts." This is also true when looking at complex technical correlations. Thus, the overall system analysis results in an added value, because it is more realistic to get the actual behavior of a technical system by using system analysis methods rather than analyzing single components or parts only.

Due to the growing demands on optimum transmission solutions—particularly with regard to drivelines or complex powertrains in automotive applications—the overall system analysis of entire drive systems becomes more and more important [1]. The "engineerlike" design of powertrains cannot be realized successfully just by studying the single parts like engines, clutches, flywheels, manual or automatic transmissions (ATs), and gear stages. According to the "spirit of the whole" (i.e., in terms of the overall system analysis), the interactions and interdependences of all drive elements due to dynamic effects have to be considered throughout the process of development and design. For this reason, powerful simulation software packages and results derived from measurements and test rig investigations are necessary to get all details and information for the best model generation and well-suited practically relevant computation methods. The interdisciplinary application of different technical areas is crucial to comply with the general technical demands. The most important technical areas concerning the analysis of drivelines are mechanics, electronics, control engineering, hydraulics, pneumatics, thermodynamics, and fluid dynamics.

Nowadays, one of the principal targets is the optimization of the dynamic and acoustic characteristics of the overall system integrated into the computer-aided classic analysis of torsional and lateral vibrations of vehicle powertrains and engines. In the future, the desired dynamic effects (mechanical vibrations, acoustic effects) are part of a design guideline to create an optimum powertrain and driveline configuration. Therefore, the *n*oise, *v*ibration, and *h*arshness (NVH) studies become a central research and development (R&D) domain. The subject "NVH analysis" is synonymous with the optimization of dynamic effects occurring in a driveline.

In the following chapters, we will focus on the analysis of the dynamic behavior of automotive drivelines in more detail. See also the standard literature to learn more on the basics of dynamics and vibrations in mechanical systems [2–5] and also the layout and computation of drive elements including the analysis of stresses and life [6]. A summary of the basic and state-of-the-art technologies of the elements in automotive drivetrains is given in Refs. [7,8].

23.2 PROCEDURE OF THE ANALYSIS OF DYNAMIC EFFECTS IN A DRIVELINE

The analysis of dynamic effects in drivelines requires a specific procedure. Due to the typical dynamic behavior of a driveline, most problems can be solved by analyzing torsional vibrations or in some cases also lateral vibrations or even the combination of both. In this chapter, we will stick to the torsional vibration analysis (TVA). There are several methods on how to handle the calculation of torsional vibrations. As most of the drive systems have a complex structure with nonlinear characteristics, a purely analytical approach is nearly impossible or difficult for the daily use in the R&D process in the automotive industry. Therefore, finite-element methods (FEMs) or multibody system (MBS) methods are commonly used for TVAs. A special and industry-proven method is the simulation in the time domain using discrete models [9,10].

The typical and recommended procedure during such a TVA is presented in Figure 23.1 [11]. This procedure will be explained in the following chapters in more detail.

23.3 TORSIONAL VIBRATION MODELS OF A DRIVELINE

In order to understand the overall system behavior of drivelines using computer-aided methods, it is necessary to create a high-quality computer model that meets practical relevance and is also easy to use. The importance of this step is often disregarded in everyday R&D life so that a more detailed model generation process is essential [1,5,9,10,12].

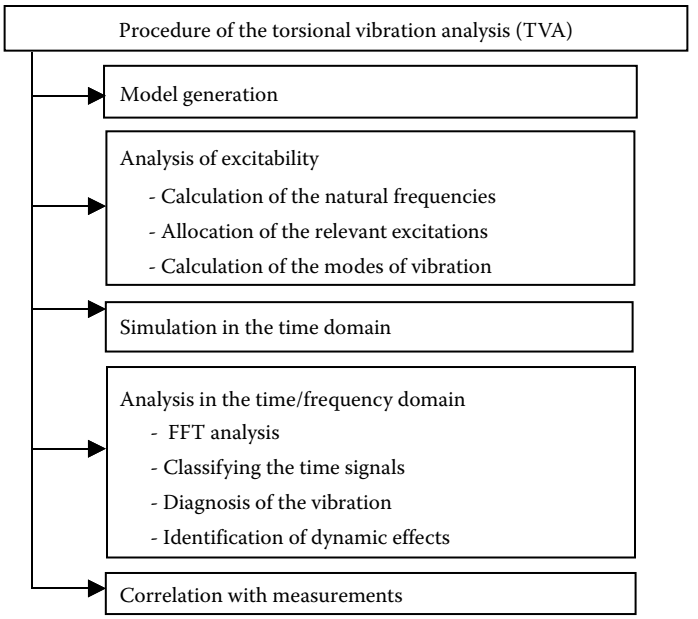


FIGURE 23.1 Procedure of the TVA.

It is recommended to generate different models according to the specific problems (e.g., for a more basic study, to solve specific problems occurring at various transmission elements like in gear stages or universal shafts or to optimize single components like dual-mass flywheel (DMF) and clutches). A uniform model for all conceivable applications is nearly impossible and uneconomical, although engineers often prefer to find such an ideal model.

The analysis of torsional vibrations in a driveline can be carried out using MBS models that are based on discrete models. Such a model consists of lumped inertias connected by stiffness elements (flexible, elastic connections). The inertias (i.e., moments of inertia) are modeled as rigid masses, and the stiffnesses (standing for shafts, gear stages, connections between two coupling halves, etc.) have no masses. Dimensions of the dampings are not necessary for the calculation of the natural frequencies and modes of the vibration; that is why an undamped system is considered first. In most of the cases, the influence of damping on the natural behavior of a torsional system is not significant (however, later we need the damping properties for the simulation!). The system structure may be straight, branched, or even intermeshed (Figure 23.2). Between two neighbor inertias, we always consider a speed ratio and a stiffness with a parallel relative damper (Figure 23.3). The dynamic behavior of a two-mass system is mathematically described in the standard literature [2–4]. The formulation is based on the following differential equation (equation of motion):

$$J\ddot{\varphi} + c\dot{\varphi} + k\varphi = T(t) \tag{23.1}$$

where

J stands for the moment of inertia
c is the relative damping coefficient
k is the torsional stiffness

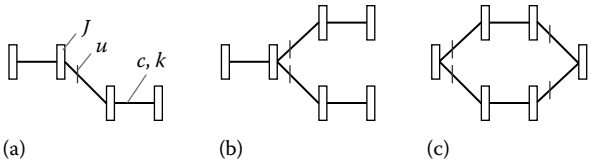


FIGURE 23.2 Structure of a discrete torsional vibration system. Examples of (a) straight, (b) branched, (c) meshed systems. *J*, rotating mass (inertia); *c*, damping; *k*, stiffness; *u*, speed ratio.

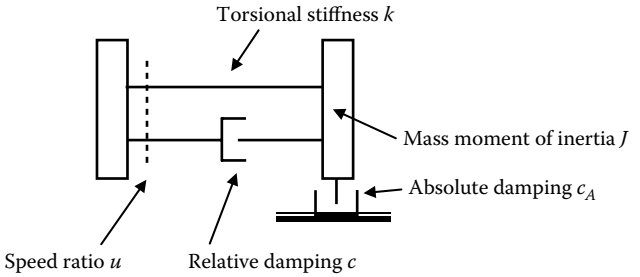


FIGURE 23.3 Basic vibration system consisting of two flexibly connected masses (inertias).

The external excitation is defined by the time-dependent torque T that may also depend on the angle φ and/or velocity $\dot{\varphi}$.

Considering the rotating "shaft element" (base element, cylinder element), where d_2 = outer diameter (O.D.), d_1 = inner diameter (I.D.), L = cylinder length, ρ = density, and G = shear modulus, the *polar mass moment of inertia* is calculated by the following general equation:

$$J_p = J = \rho I_p L \ [\text{kg m}^2] \tag{23.2}$$

The axial moment of inertia—as mentioned in Equation 23.2—can be calculated as follows:

1. *Solid shaft*

$$I_p = \frac{\pi}{32} d^4 \ [\text{m}^4] \tag{23.3}$$

2. *Hollow shaft*

$$I_p = \frac{\pi}{32} (d_2^4 - d_1^4) \ [\text{m}^4] \tag{23.4}$$

The *torsional stiffness* is

$$k = \frac{GI_p}{L} \ [\text{N-m/rad}] \tag{23.5}$$

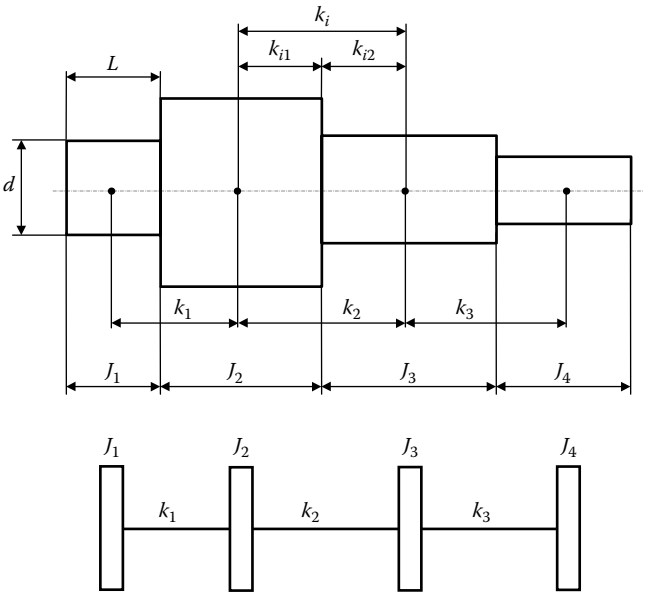


FIGURE 23.4 Multiple stepped shaft with indicated sections as lumped inertias and stiffnesses (d = diameter, L = length, J = mass moment of inertia, k = torsional stiffness).

The formulas show that either the moment of inertia or the stiffness is proportional to the fourth power of the diameter. Considering the length of the cylinder, the moment of inertia increases linearly, while the torsional stiffness is inversely proportional to the length of the considered element. Therefore, it is important for the model validation to check the accuracy of the diameters.

For the calculation of the moments of inertia and the stiffnesses, the shear modulus ($G = 8 \cdot 10^{10}$ N/m^2 for steel) and the material density ($\rho = 7650$ kg/m^3 for steel) must be considered.

If a torsional vibration model should reflect a multiple stepped shaft (Figure 23.4), we have to model every "significant part" of the shaft either as rigid inertia or as stiffness element. The splitting of the shaft into "significant parts" means that the real geometries are split into mass concentrations (i.e., mass moments of inertia) or in elasticities (i.e., torsional stiffnesses). Often the cylindrical elements of such a stepped shaft are modeled as mass moments of inertia, whereas the connections in between become representative torsional stiffnesses. It is important to note that the lining of "partial stiffnesses" leads to an "*elastic series connection*." The total stiffness is always less than the smallest partial stiffness due to the reciprocal accumulation of the stiffnesses (Equation 23.6).

Moments of inertia always refer to the center of gravity of the shaft sections. If necessary, a shaft section combines several smaller sections to a larger single inertia element. In this case, the moments of inertia are calculated in the form of series connections between adjacent inertia elements. The multiple stepped shaft in our example shows several shaft sections that correlate with the discrete structure of the torsional vibration model.

The series connection of multiple stiffnesses (general formulation) is

$$k_{total} = \frac{1}{\sum_{\nu} 1/k_\nu} \tag{23.6}$$

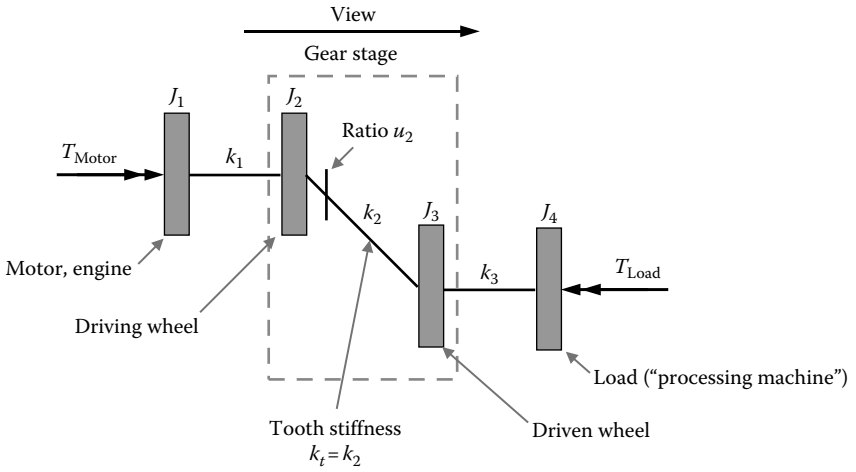


FIGURE 23.5 Vibration model showing details of the properties used to define gear stages.

For the calculation of the *tooth stiffness* k_t (this implies the stiffness between two gear wheels or pinions), the following basic approach for the determination of the substitute stiffness can be used [9]:

$$k_t = k_{spec} \frac{b}{\cos\beta} r_0^2 \,[\text{Nm/rad}] \tag{23.7}$$

where

k_{spec} is the specific tooth stiffness, for example, 10–20 [N/mm/μm]
b is the tooth width [mm]
β is the helix angle [degree]
r_0 is the pitch circle radius (related to the subsequent pinion) [mm]

The recommended values for k_{spec} are based on the literature [13,14]. A generally accepted figure for the mesh stiffness k_{spec} of normal teeth typically for spur gears is 1.4×10^{10} N/m/m [14,15]. However, in order to get the best values for the relevant tooth stiffnesses, we have to take into account the significant effects of the tooth design that can be numerically evaluated by modern FEM models. Different manufacturing methods produce different root shapes and affect stiffness, but the main variations arise from variation of pressure angle or undercutting.

A typical vibration model assuming a tooth stiffness k_t as "torsional stiffness" is presented in Figure 23.5. For the calculation of the tooth stiffness, the pitch circle radius must be applied to the driven wheel with the inertia J_3.

The model size depends on the complexity of the engineering targets. The maximum natural frequency in the computer model should always be greater than the maximum most significant excitation and acoustic frequencies [11].

In order to reduce a model in terms of a reduction of inertia and stiffness stations and therefore a minimization of the maximum natural frequency, a special method based on Rivin's approach [5,9,12] should be used. Reduction guidelines should be noticed in general [11].

23.4 ANALYSIS OF EXCITABILITY

23.4.1 Basics

For the interpretation of the dynamic behavior of a driveline, the "analysis of excitability" is necessary [11]. The first step of the analysis of the dynamic behavior of a drive system comprises the analysis of the natural behavior (in terms of natural frequencies or eigenvalues and

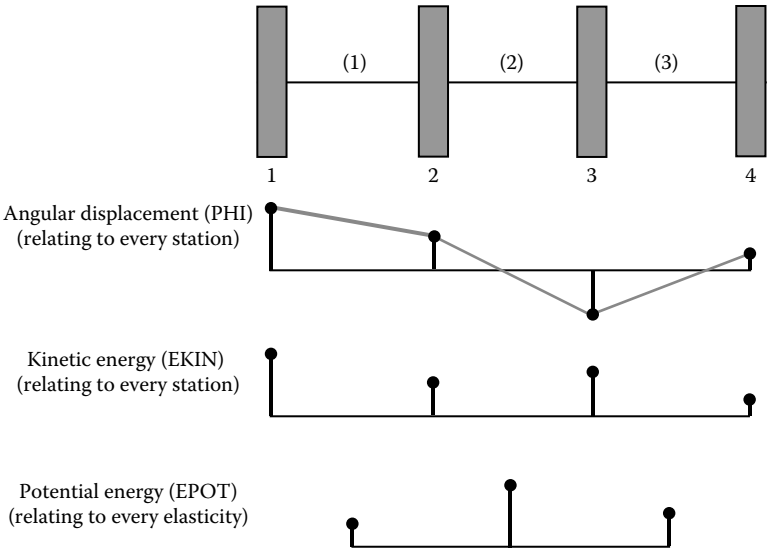


FIGURE 23.6 Presentation of a vibration mode referring to a specific natural frequency.

the corresponding modes or vibration modes or eigenmodes). The system natural behavior is the "portrait" of a dynamic system and therefore of significant interest for the basic understanding of a dynamic analysis. The natural frequencies and modes are usually calculated for the undamped and linearized system.

As an example, Figure 23.6 shows the modes of vibration for one specific natural frequency of a straight four-mass system (four moments of inertia, three stiffnesses) visualized as the normalized values of the angular displacement (PHI), the kinetic energy (EKIN) of every rotating inertia, and the potential energy (EPOT) for every torsional elasticity.

In order to interpret the modes, the following aspects should be noticed:

1. The EKIN amplitudes represent the most important stations (inertias) that might be sensitive in case of external excitations (e.g., engine, flywheel, or any other large inertias) at this station.
2. The EPOT amplitudes represent the most important elasticities (torsional stiffnesses) that might indicate a critical element in resonance, if an excitation does exist with the corresponding excitation frequency and a considerable dynamic amplitude.
3. The combined interpretation of the EPOT and EKIN modes shows how far the drive system experiences a resonance-type amplification in a specific driveline element. This is expressed by "excitability." It must be noticed that a quantitative statement can only be made by simulation methods.

23.4.2 Sensitivity of the Natural Behavior

For the global, superior system understanding as well as for the system identification, the knowledge of the sensitivity of the system in case of the analysis of excitability is crucial. With the help of the statement about the sensitivity of a system, the question concerning the sensitivity of the variation of particular parameters and also the sensitivity due to tolerances and uncertainties of data can be answered, too. That offers not only the opportunity but also the danger to influence the dynamic characteristics by the specific variation of particular system parameters (e.g., stiffnesses and inertias). Further information can be taken from the literature [9,16]. Figure 23.7 shows a two-parametric evaluation of the

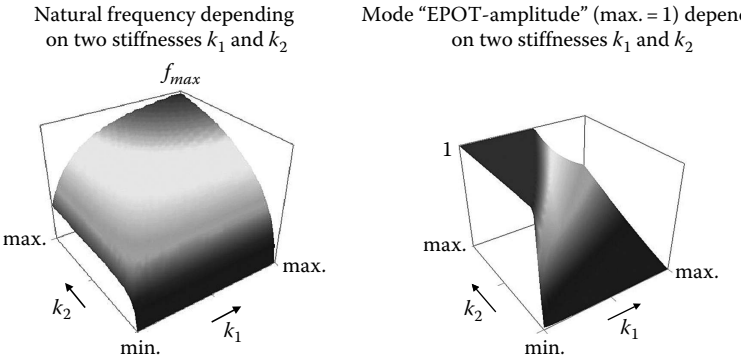


FIGURE 23.7 Analysis of the sensitivity of the natural driveline behavior by the variation of characteristic elasticities (k_1 = stiffness for clutch/damping plate, k_2 = stiffness for side shafts of a rear-wheel-driven car); the amplitudes of the normalized modes are visualized by the gray shades.

system behavior of a vehicle driveline with reference to a specific mode at a selected natural frequency. Additionally, measurements can be rated and become understandable with the help of the sensitivity analysis of the natural behavior.

23.4.3 Resonance Diagram

Usually resonances are graphically presented in a so-called Campbell diagram (*resonance diagram*). This diagram presents the frequencies (=speed-dependent excitation frequencies for every order and also the speed-independent natural frequencies) versus speed (e.g., engine speed) and may visualize possible resonance speeds. The order lines correspond to the significant harmonic excitations.

The following table assigns possible excitation sources to their typical excitation frequencies (Table 23.1). The speed n [rpm] is always the speed of the corresponding drive element that produces the excitation.

There are several types of excitations. One excitation line might correspond to the second order of a combustion engine (e.g., four-stroke four-cylinder engine); another excitation line refers to the excitation of an aggregate [e.g., compressor and pump]. Furthermore, excitation lines are also caused by so-called parametric excitations as found in gear stages due to the fluctuating tooth stiffnesses (mesh frequencies).

TABLE 23.1
Excitation Source and Corresponding Excitation Frequencies

	Excitation Source	Excitation Frequency Depending on the Frequency $f = n/60$ [Hz] Where n = Speed [rpm]
1	Clutch (+backlash), compressor, turbine; excitation due to unbalances or eccentricities	**$1 \times f$ or $2 \times f$**
2	Universal shaft, slider crank	**$2 \times f$**
3	Gear drive	Gear mesh frequency
	z_G = number of teeth	**$1, 2, 3 \times z_G \times f$**
4	Pump, fan, ventilator	Blade-passing frequency
	z_P = number of blades	$z_P \times f$
5	Compressor, pump, two-stroke combustion engine	**$1, 2, 3, \ldots, k \times f$**
6	Four-stroke combustion engine	**$0.5, 1.0, 1.5, 2.0, \ldots, k \times f$**

Speed-proportional excitations of electrical drives, which depend on the motor characteristics, can also be visualized as typical excitation lines within a resonance diagram. As unbalances and eccentricities often occur in drivelines, the first and second excitation orders referring to every speed of the shafts can always be noticed.

Since the natural behavior is usually calculated according to a linear torsional vibration system, the time- and speed-independent natural frequencies can be displayed as horizontal lines. Intersections between the natural frequency lines and the order lines can lead to resonances. A resonance, which relates to a specific station of the driveline (e.g., a stiffness station in case of a torque-response signal), only exists as important magnitude (e.g., amplified torque response), if the following items are effective all together:

1. Excitation and natural frequency coincides.
2. Excitation amplitudes are significant.
3. Corresponding vibration mode shows significant amplitudes at the station, which is analyzed.

Large resonance peaks can lead to damages inside the driveline. A stress and life analysis must show that the current excitations are within the allowable limits [6]. Smaller resonance peaks do not affect the durability, but they can lead to serious noise problems inside the driveline or to disturbing vibrations (quality issues). Therefore, even small mode energies (i.e., small amplitudes in the vibration mode) may cause acoustic problems. This fact plays a major role in the NVH analysis. Typical resonance diagrams are presented in Figure 23.8.

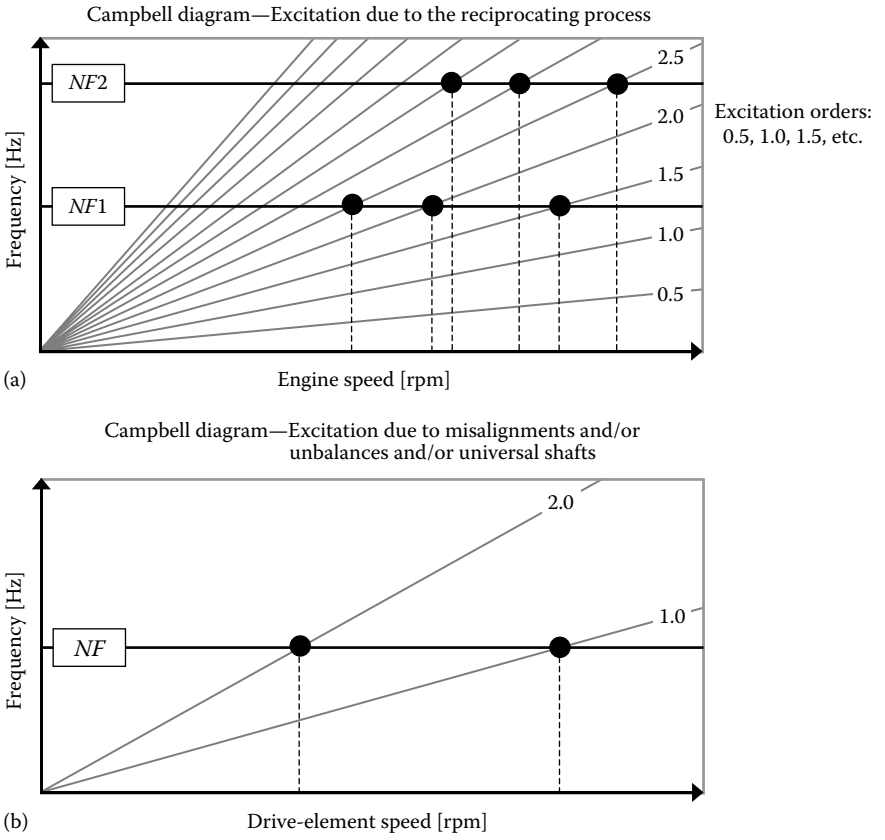


FIGURE 23.8 Typical resonance diagrams with highlighted resonance points (=dark circles): (a) engine excitation, (b) misalignments and/or unbalances and/or universal shafts.

(*continued*)

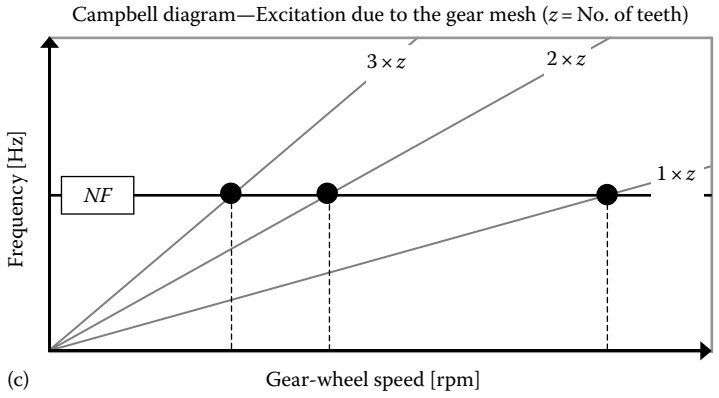
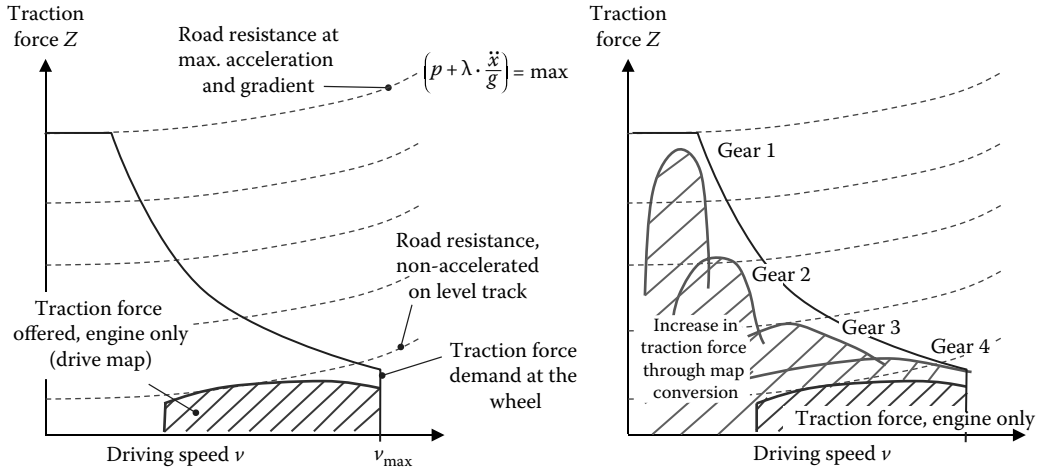


FIGURE 23.8 (continued) Typical resonance diagrams with highlighted resonance points (=dark circles): (c) gear mesh (dynamic tooth stiffness). *NF* = natural frequency, z = number of teeth.

23.5 COMPONENTS OF A DRIVELINE

The basic functions of the drivetrain include the support of the vehicle optimized in terms of fuel consumption, emissions, noise, comfort, and performance. It focuses on the conversion of the primary engine maps into secondary wheel maps (Figure 23.9). Starting devices and transmissions, which have to meet specific demands, are used in conventional drivetrains to fulfill these functions. General descriptions of the components in an automotive drivetrain are given in Ref. [8].

Figure 23.10 exemplifies a standard drivetrain used in many vehicles. A longitudinally mounted combustion engine with flange-connected starting device and transmission is typical of this drivetrain design. The power is transferred by a propshaft from the transmission output to the rear axle gear with integrated differential. The differential outputs are connected to the wheels by the drive shafts. The differential is required for wheel speed compensation of the driven axle in case of cornering.



Primary function of the transmission: conversion and adjustment of the engine map to the demand map.

FIGURE 23.9 Adjustment of the power map of a combustion engine to the traction force demand at the wheel by conversion of the map due to a multispeed transmission. Acceleration and uphill or downhill driving would be limited without conversion.

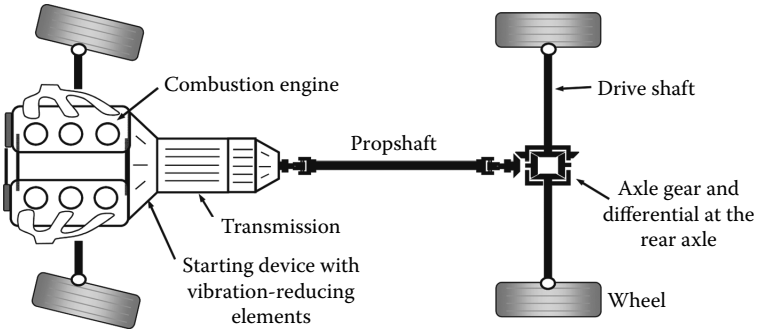


FIGURE 23.10 Components of a standard drivetrain: The power in the conventional drivetrain is transferred from combustion engine to the transmission by the starting device and to the wheels by shafts and axle gear.

Besides this design, front longitudinal or front transverse installations are common. For the first one, the propshaft is not required since the power is transferred to the drive shafts of the front axle by an axle gear and differential integrated in the transmission. The latter one includes a combustion engine mounted transversely to the longitudinal vehicle axle and usually also a transmission with integrated axle gear and differential. All drivetrains have specific requirements on the transmission design (see Section 23.5.3).

The mechanical power required to drive the vehicle is generated by the combustion engine and is available as speed and torque at the transmission input shaft. It is typical of modern internal combustion piston engines that they can only produce torque when the idle speed is reached and that they have a maximum speed. Other important characteristics are the fluctuation of torque due to single ignition of each piston and the resulting irregular crankshaft speed. To reduce the irregularity, flywheels and torsional vibration dampers are installed in modern drivetrains (see Section 23.5.2).

In conventional drivetrains, starting devices are used to bridge the speed gap resulting from the necessary minimum engine speed. The most common types of these devices include mechanical friction clutches and Trilok converters (cf. Section 23.5.2.2). They are mounted in front of the transmission. The transmission is the most important part involved in the conversion of the maps. The ratio of the transmission as a speed torque converter can generally be adjusted in different gears. The most important transmission designs in serial production will be introduced in Section 23.5.3. The ratio of the axle gear is usually constant. The overall ratio of the drivetrain is determined by the product of both the transmission ratio and the axle gear ratio. The overall ratio normally ranges from 2 (highest gear) to 16 (starting gear):

$$i_{Overall} = i_{Transmission} \cdot i_{FinalDrive} \tag{23.8}$$

Propshafts within the drivetrain transmit rotary power and are able to compensate spatial displacement of components, for example, caused by wheel angle while cornering. The wheel is finally responsible for the conversion of the rotary drivetrain movements and torques into translational movements and forces.

As a matter of fact, the combination of all mentioned rotating drive elements in a vehicle driveline is critical for the overall dynamic sensitivity and also NVH characteristics of the total rotating system [17]. Therefore, it is a must to analyze the dynamic behavior of this system as a whole and to understand the detailed characteristics and importance of every single drive component.

23.5.1 Engine

Combustion engines are reciprocating machines—as typically used in vehicle drives—and are usually working according to the four-stroke process. As the name implies, the operation of a four-stroke internal combustion engine comprises four basic steps that repeat with every two revolutions (720°) of the engine (Figure 23.11). The typical dynamic engine excitation is

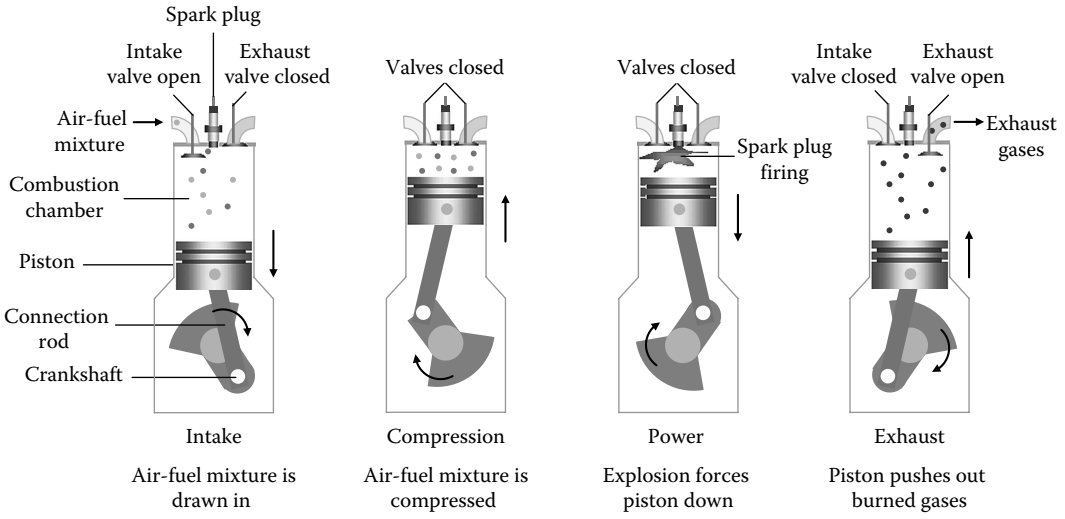


FIGURE 23.11 Four-stroke cycle of a reciprocating combustion engine in a vehicle.

described by tangential force curves or recalculated as torque-angle functions, mostly represented by Fourier coefficients depending on the process characteristics (gasoline, Diesel) as presented in Figure 23.12 [18,19].

The rotational irregularity at the crankshaft is caused by the combustion characteristics of combustion engines. Only when the fuel is burnt, a drive force is generated that produces torque at the crankshaft. For a four-cylinder four-stroke engine, there are thus two force peaks per crankshaft rotation due to two-plug ignition. Irregularities of the second engine order can thus be observed at the crankshaft. This reduces the service life of the drivetrain components due to continuously changing torque load on the one hand, and on the other, vibrations in the drivetrain are induced that are perceived as influencing the comfort negatively.

23.5.2 Flywheel and Clutch

23.5.2.1 Torsional Vibration Damping at the Transmission Input

A combination of increase in inertia and vibration damping is usually applied in the drivetrain to reduce the effects of the engine-related (primary side) vibrations on the residual drivetrain. In the simplest case, a flywheel is used to increase the rotary inertia of the crankshaft, so the torque fluctuation does not influence the speed very much. This additional mass, however, affects performance and fuel consumption negatively, since inertia to be accelerated in driving maneuvers increases significantly particularly in lower gears. A clutch–spring–damper system is therefore integrated in the clutch disk (Figure 23.13). Suitable spring stiffness leads to uncritical vibration response of the drivetrain on the transmission input side, with its natural angular frequency ν being reduced in a way that it results in an isolation frequency Ω below the double idle frequency of the combustion engine (Figure 23.15a). The vibration irregularity in normal vehicle operation on the primary side is damped considerably due to the resulting damping characteristics, so the load on the transmission input is reduced. Vibration amplification only occurs when starting the engine.

Conventional clutch–spring–damper systems, however, can only damp rotational fluctuations up to a certain amplitude. Furthermore, good damping characteristics at low excitation frequencies, that is, low engine speeds, are achieved by low stiffness of the spring elements, which leads to perceptible load-alteration effects that are thus evaluated as negative. But modern car engines are characterized by higher torques even at low speeds as well as reduced idle speed and fewer pistons (downsizing), resulting from the wish for reduced fuel consumption and emissions. This in turn leads to increasing

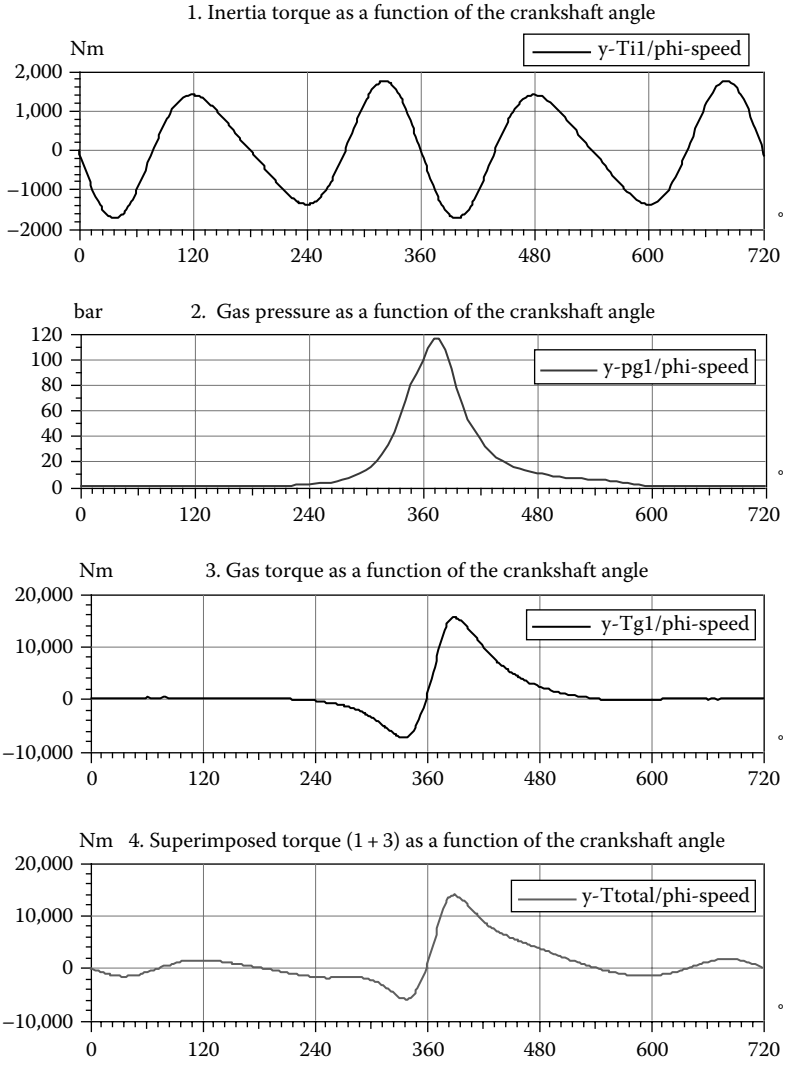


FIGURE 23.12 Typical presentation of engine excitation torques.

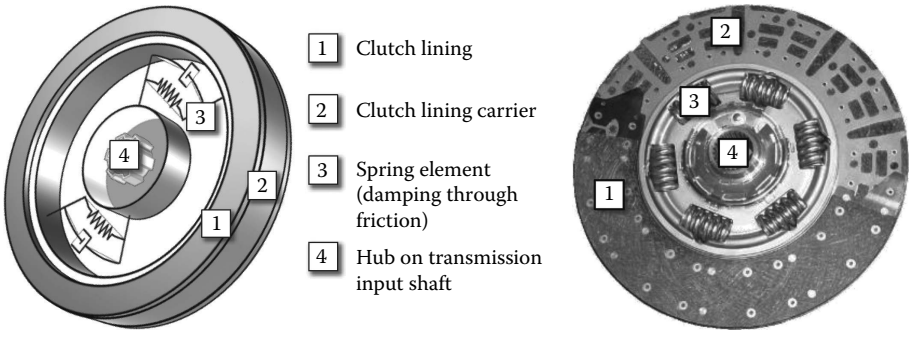


FIGURE 23.13 Conventional clutch–spring–damper system. There are tangential spring packs between the inner part with connecting hub to transmission input shaft and the friction lining carrier. The damping effect is caused by friction.

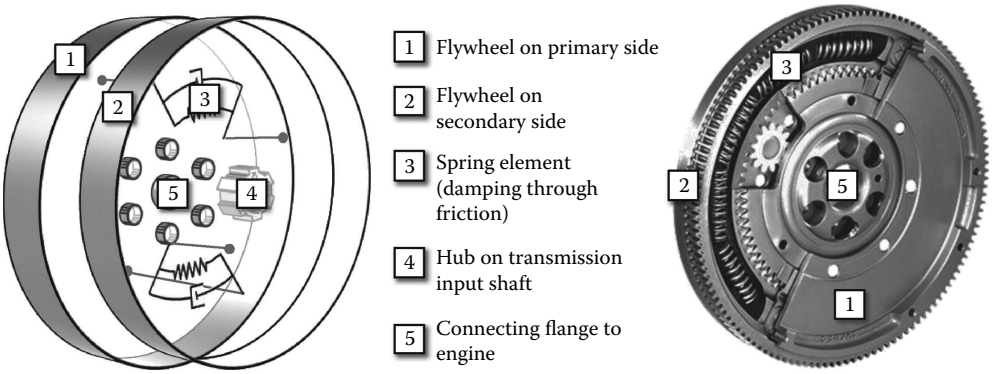


FIGURE 23.14 DMF: Flywheels on the primary and secondary sides are connected by a spring pack. Flange and hub connect them to engine and transmission.

rotational irregularity especially in the critical range of low frequencies. For this reason, alternative technologies are required to protect the transmission input efficiently from torsional vibrations. Established technologies include the so-called DMF (Figure 23.14). Here, the conventional flywheel (single-mass flywheel) of the crankshaft is divided into a primary and secondary flywheel, and a (multistep) tangential spring system is placed between them. The vibration system at the transmission input is thus changed in a way that additional vibration modes are generated and the characteristic amplification function of rotational irregularities of the engine is changed in a positive way: the isolation frequency shifts toward lower speeds below the engine idle speed due to the less dominant natural frequency (Figure 23.15b). The main disadvantage of this technology is the high cost for a DMF.

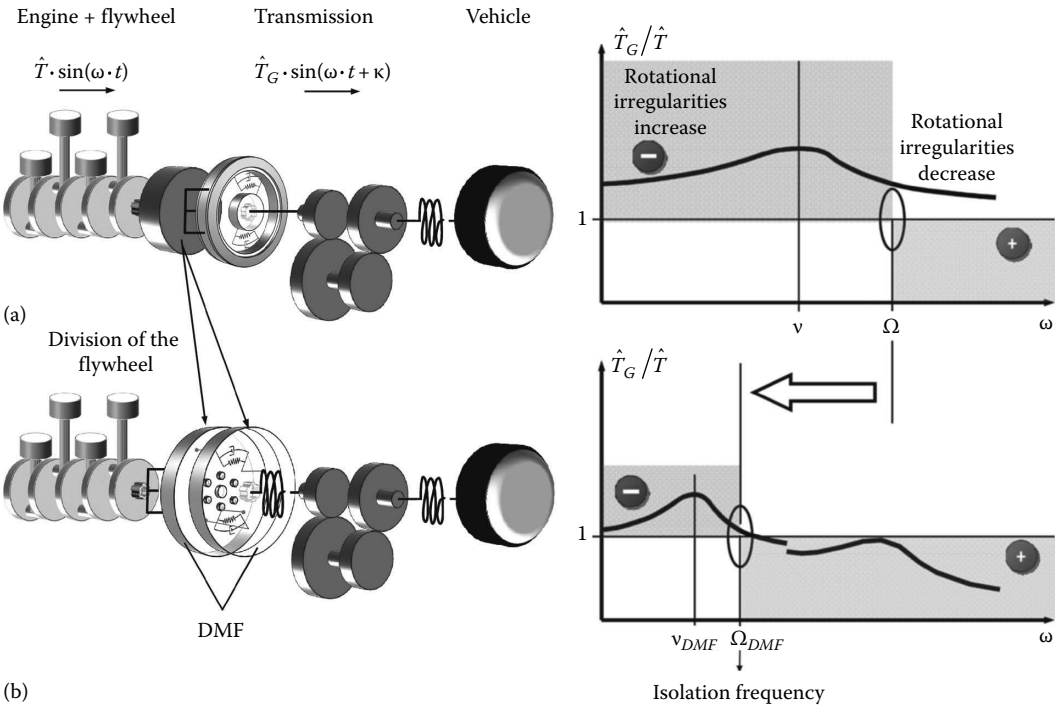


FIGURE 23.15 Vibration absorption at transmission input by conventional vibration dampers (a) and DMF (b).

This is the reason why researchers look for alternative technologies that guarantee good vibration damping with only minor additional consumption (e.g., centrifugal pendulum).

Extended details are found in Refs. [20–22].

23.5.2.2 Starting Devices

As described earlier, even the most modern combustion engines are able to provide power at the crankshaft when reaching a certain speed. Below this limit, the torque generated by the pistons is inadequate to overcome friction loss. Therefore, it is necessary to include a starting device in the drivetrain that helps to bridge the speed gap. What the nowadays often used starting device clutch and torque converter have in common is the power loss that occurs in the form of heat that has to be stored or emitted when starting. In general, the requirements on starting devices can be summarized as follows:

- Separating drivetrain from running engine if the vehicle is at standstill or during gear change
- Bridging speed when starting
- Vibration isolation
- "Overload protection" in case of blocked engine

23.5.2.3 Mechanical Clutches

Mechanical friction clutches, either as dry clutches for manual or Dual-clutch transmissions (DCTs) or as wet multidisk clutches with oil cooling in DCTs, are not torque converters. This means that the speeds for both halves of the clutch can differ, but the torque for both parts is identical.

Figure 23.16 represents the layout of a dry mechanical friction clutch. It usually consists of a clutch disk with friction lining between a pressure plate and the flywheel or the secondary flywheel of a DMF. The clutch disk is connected to the transmission input shaft so as to prevent rotation, and spring packs can be integrated in case of a conventional clutch–spring–damper system. Flywheel and pressure plate rotate together with the engine speed. If the clutch is not actuated, the pressure plate is pressed to the clutch disk by a disk spring that induces normal force. This force dictates the

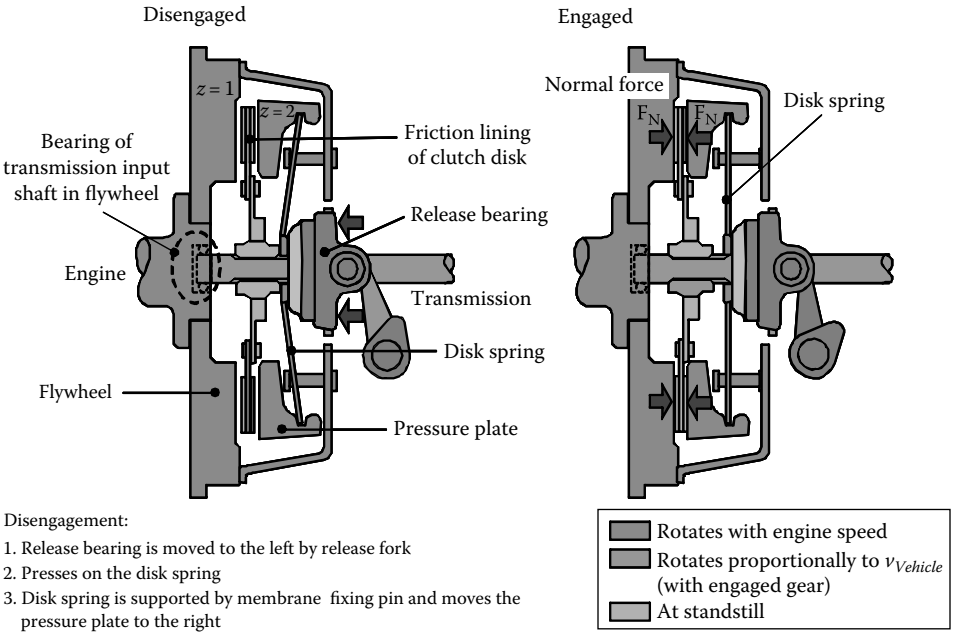


FIGURE 23.16 Structure and function of a mechanical friction clutch with disk spring.

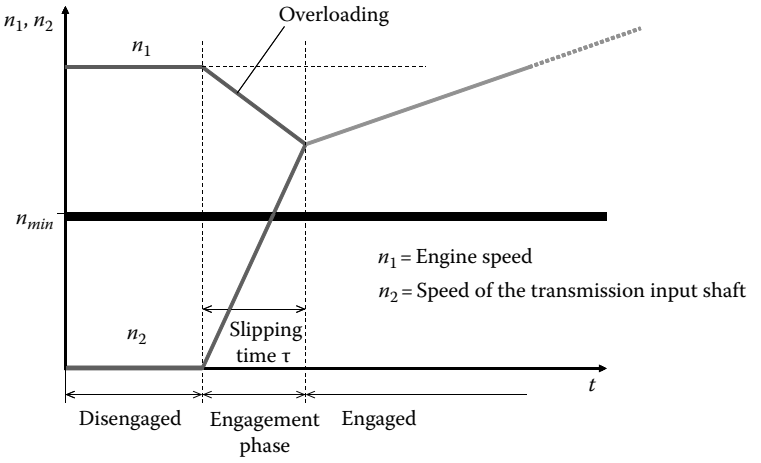


FIGURE 23.17 Speed curves of engine and transmission input during starting. During the engagement phase, the clutch slips and thus transmits torque. If the accelerator is not increased, the engine speed decreases slightly until the transmission input speed reaches the level of the engine speed.

transmittable clutch torque. During disengaging, the load on the disk spring is released by a release bearing and the pressure plate is separated from the clutch disk. The spring characteristics affect the force in the clutch pedal significantly. In case of an automated manual transmission (AMT), the actuators move the release bearing.

During engaging the clutch slips, that is, both parts of the clutch rotate at different speeds with normal force and transmitted torque. During the start-up process as illustrated in Figure 23.17, the clutch slips in the "engagement phase." The clutch transmits torque from the engine to the transmission and thus to the wheels. The vehicle is accelerated and the transmission input speed increases, while the engine speed is reduced—assuming the driver does not increase the accelerator. As soon as transmission input speed and engine speed are identical, the state of the clutch changes from friction to adhesion and the engaging process is completed.

During the engagement phase, friction power loss occurs in the clutch in the form of heat input to the clutch lining. Thus, the clutch heats up considerably. The heating depends on transmitted torque, vehicle mass, and starting gear ratio. For this reason, not all heavy and powerful vehicles can be equipped with dry friction clutches. Wet shift elements or the so-called hydrodynamic torque converter is therefore used.

23.5.2.4 Hydrodynamic Torque Converter

Hydrodynamic torque converters are based on the principle of hydrodynamic clutches. They consist of a pump that is connected to the engine and a turbine connected to the transmission input. Oil circulates between both of them (Figure 23.18). In case of the torque converter, a so-called stator is added to this type of clutch. Up to a certain speed ratio of turbine and pump, the stator diverts the flow of the oil, so the torque at the turbine is increased. The range where the speed ratio $\nu = n_{Turbine}/n_{Pump}$ is smaller than in the so-called coupling point is called converter range. The smaller the speed ratio, the higher the so-called torque increase $\mu = T_{Turbine}/T_{Pump}$ (Figure 23.19). The reason for this is flow diversion at the stator, which causes torque at the stator that is supported by the sprag clutch. This stator torque is also available at the turbine in addition to the torque from the pump. The hydrodynamic two-phase converter, also named Trilok converter after the inventor, is therefore not only a speed converter but a simple hydraulic transmission in the form of a speed torque converter. At the coupling point and beyond, the flow at the stator is changed in a way that it rotates over the sprag clutch and the torque is not increased.

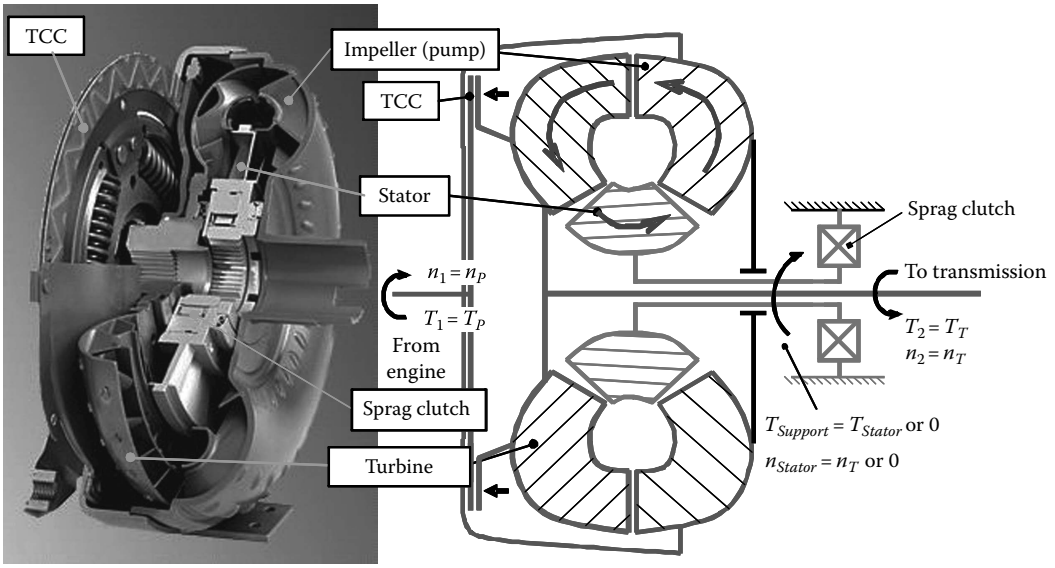


FIGURE 23.18 Construction of a hydrodynamic two-phase converter (Trilok converter).

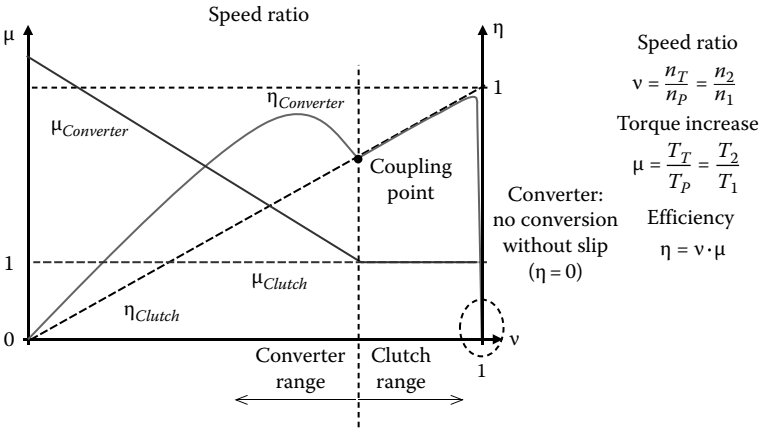


FIGURE 23.19 Torque ratios and efficiency of mechanical friction clutches and hydrodynamic two-phase converter.

While the clutch slip determines the efficiency of mechanical clutches, higher efficiency during starting can be achieved with a hydrodynamic converter due to the increase in torque (Figure 23.19). An important disadvantage of the converter is that power can only be transmitted if there is a speed difference between pump and turbine. This results in reduced efficiency during constant driving. For this reason, nowadays a so-called torque converter lockup clutch (TCC) is used, which connects pump and turbine mechanically. Furthermore, this mechanical connection reduces the so-called rubber band effect, which describes the roaring of the engine in case of load reversal and the sluggish response of the vehicle. But to use the good damping characteristics of the converter especially during gear change, the TCCs are used with controlled slip in some operating states that are critical in terms of vibration.

Due to their design, the hydrodynamic converters allow the transmission of high torques. The good damping characteristics as well as converter increase lead to high driving comfort, which is the reason why this type of starting devices is mainly used in high-quality vehicles.

Extended details are found in Refs. [23,24].

23.5.3 Manual and Automatic Transmission

The transmission does not only convert speed but also torque. This means that speed and torque at the transmission input differ from those at the output. The power, however, remains the same, except for losses (efficiency).

There is at least one ratio resulting from the relation of the speeds at transmission input and output:

$$i_{Transmission} = \frac{n_{Transmission\ Input}}{n_{Transmission\ Output}} \tag{23.9}$$

According to the type of transmission, one gear (axle gear), several gears (manual transmission = MT), or an infinite number of gears (continuously variable transmission = CVT) are possible, with speed-reducing ratio ($i > 1$) as well as speed-increasing ratio ($i < 1$, so-called overdrive gears).

Taking the transmission efficiency

$$\eta_{Transmission} = \frac{p_{Transmission\ Output}}{p_{Transmission\ Input}} = 1 - \frac{p_{Transmission\ Loss}}{p_{Transmission\ Input}} \tag{23.10}$$

into account, the following relation between the transmission input torque and the output torque $T_{Transmission\ Output}$ applies:

$$T_{Transmission\ Output} = T_{Transmission\ Input}\, i_{Transmission}\, \eta_{Transmission} \tag{23.11}$$

The following criteria can be used to classify vehicle transmissions:

- *Degree of automation*: There are "MTs" (cf. Section 23.5.3.1) as well as "semiautomatic and fully ATs" (AMT, CVT, DCT, AT; cf. Sections 23.5.3.2 through 23.5.3.5), with automatic starting and shifting as well as automatic selection of shift points [25]. Principally, the efficiency of MTs is higher than of ATs, which always require additional energy to operate the actuators. This energy cannot be used to drive the vehicle.
- *Type of power transmission*: The drive power can be transmitted in different ways. It is transmitted mechanically in conventional transmissions, and fluid-mechanically transmitted only in the hydrodynamic converter. In modern hybrid drivetrains, there are so-called electric power split transmissions with mechanical as well as electric power transmission.
- *Strategy of ratio adjustment*: In this context, stepped ATs with discrete steps and "CVTs" are differentiated (CVT; cf. Section 23.5.3.3). The ratio of the CVT can be selected freely within the limits given by the design.
- *Design*: According to the drivetrain layout, different transmission concepts are required. Transmission input and output of so-called *inline transmissions* are arranged coaxially. Spur gear transmissions (MT, AMT, DCT) therefore have to have at least two ratio steps (Figure 23.20a): in general, there is a constant transfer from the input to the so-called countershaft and gear ratio to the output shaft. Planetary transmissions (AT, cf. Section 23.5.3.5) comply with this. Inline transmissions are especially used for standard design drivetrains (Figure 23.10). This design allows for a direct gear with a ratio of and high efficiency with the torque being directly transmitted from input to output shaft. If the engine is mounted longitudinally and the axle of the vehicle is driven, where the engine is mounted, so-called *transaxle transmissions* are used primarily (Figure 23.20b). This type of transmission combines transmission ratio and axle drive and allows for the required rotation of the pivot axle by 90°. For transversally mounted engines, *transverse transmissions* are used (Figure 23.20c). They have two or more parallel shafts, but the output is not coaxial to the input. Axle gear and axle differential are generally integrated in the transmission.

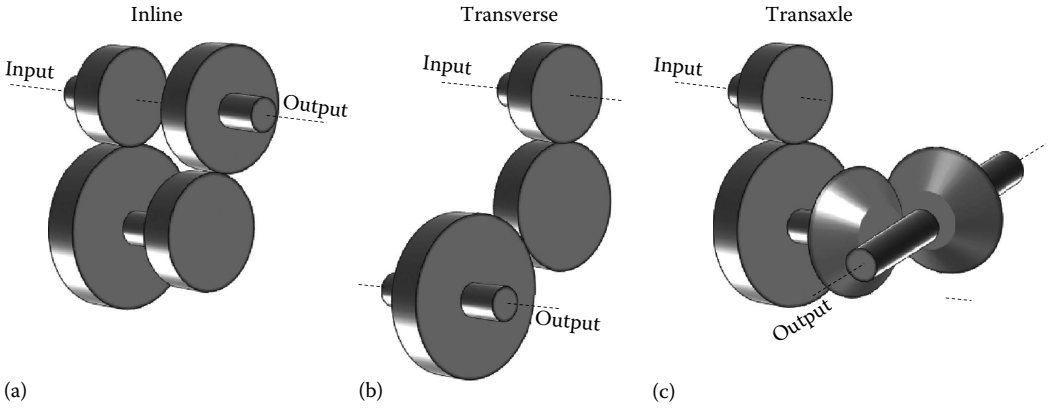


FIGURE 23.20 Transmission designs for different drivetrains. (a) Input and output shafts of standard drivetrains are arranged coaxially. (b) The axle positions of transmissions for transversally mounted engines are parallel but different, usually with an integrated axle gear. (c) Transaxle transmissions are used for longitudinally mounted engines and combine primary transmission and axle gear but with a 90° rotation around the pivot axis.

- *Interruption of traction*: In transmissions that only have one clutch as a starting device (MT, AMT), the clutch usually has to be opened during shifting since the gear change requires an unloaded transmission. When the clutch is opened, power cannot be transferred to the wheel, so the traction is interrupted. During acceleration, this results in a considerable longitudinal acceleration interruption (cf. Figure 23.21). Transmissions with more than one clutch (DCTs and ATs), however, can change gears without interruption of traction since the torque of a (disengaging) clutch is transmitted to another (engaging) one. The torque is transmitted to both clutches during gear changes, so the driver ideally does not recognize the change.

In the following, the most important transmission designs will be described.

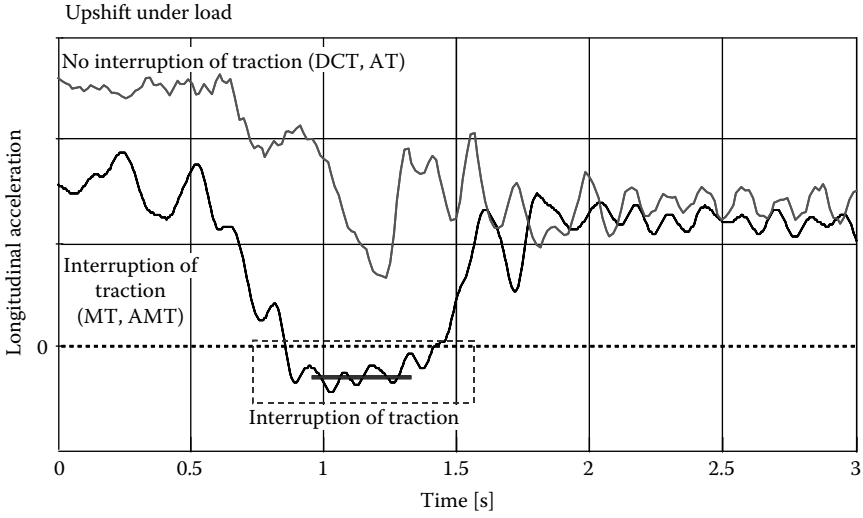


FIGURE 23.21 Curves of longitudinal acceleration during shifting under load with and without interruption of traction. In case of transmissions with traction interruption (MT, AMT), the longitudinal acceleration decreases significantly and is partially negative, while the changes in longitudinal acceleration are hardly recognizable when the traction is not interrupted.

23.5.3.1 Manual Transmission

MTs are the best accepted type of transmissions in Europe as well as in many important markets worldwide due to their low manufacturing costs and high efficiency. MTs are spur gear transmissions. The gear pairs for each gear are arranged on at least two parallel shafts, each pair consisting of a fixed gear and an idler gear. The fixed gear is fixed to a shaft, while the idler gear can rotate freely on the other shaft. Only if a gear is engaged, idler gear and shaft are connected by a shift unit (often synchronization units).

Figure 23.22a illustrates the structure of an MT for inline installations: the torque is transmitted constantly from the input (left) to the countershaft. According to the engaged gear, the torque is transmitted to the output shaft coaxially to the input shaft and from the output shaft to the axle gear. The fifth gear as a so-called direct gear is special since the power is directly transmitted from the input to the output shaft. An additional intermediate gear is required for the reverse gear, which causes the reversal of direction of rotation. Figure 23.22b shows the structure of a transverse transmission. As opposed to the inline transmission, the gear-dependent torque is directly transmitted from input shaft to countershaft, which is also the output shaft. The torque is then transferred to the integrated axle gear. Since the axial space—particularly in case of transversally mounted engines—is limited, small transmissions are required. Therefore, so-called multi shaft concepts are often used, consisting of several countershafts (partial shafts). The torque is transmitted from the input shaft to the countershafts and then to the differential (Figure 23.23).

There are many modern approaches to optimize the gear design and gear shift concepts [26–29].

23.5.3.2 Automated Manual Transmission

AMTs are basically of the same design as MTs, but the driver actions "operating the clutch" and "changing gears" are replaced by automatic operations through electric or hydraulic actuators. This is controlled by an electronic transmission control unit with complex control and communication structure. The driver is therefore relieved of these tasks, and fuel consumption can be reduced through an optimized selection of shift points [30,31].

The AMT is especially used for sports cars due to very short shifting times and high efficiency—the trend, however, is toward DCT instead of AMTs. The primary advantage of the AMT

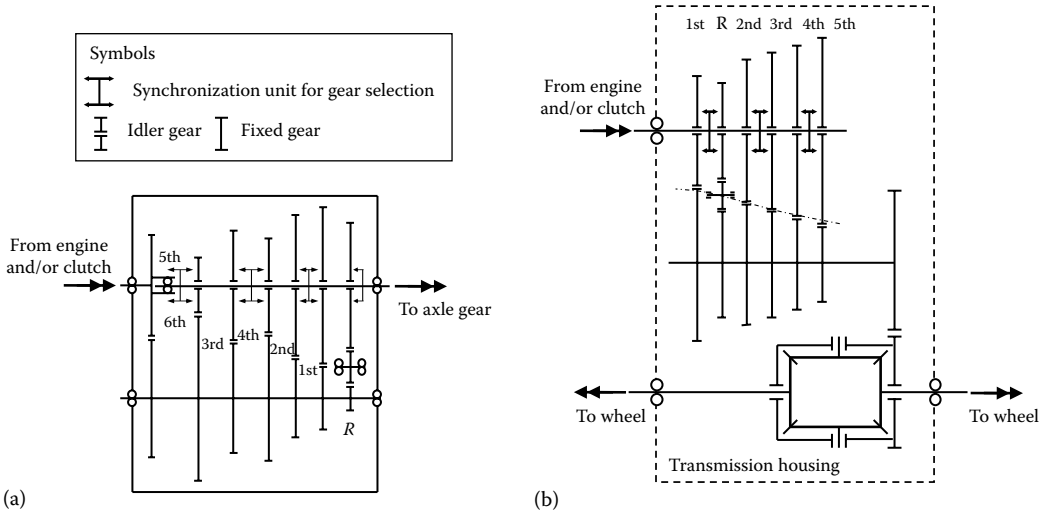


FIGURE 23.22 Layout of an MT for (a) inline use (e.g., standard drive) and transverse use (e.g., front-transverse). In case of (b), axle gear and differential are already integrated.

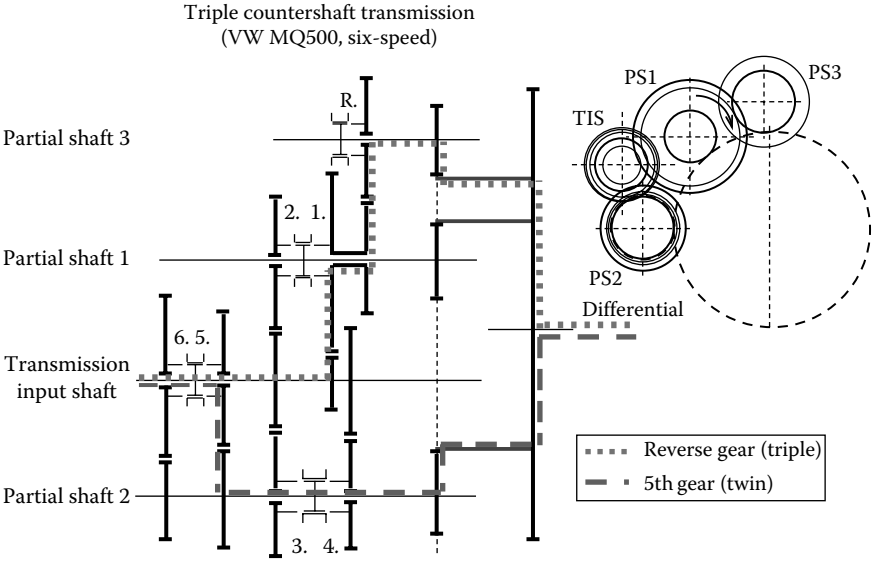


FIGURE 23.23 Six-speed triple-countershaft transmission MQ500 by Volkswagen with several partial shafts that are connected through the big differential gear.

in connection with small vehicles (lower price class) is the cost. It is quite uncomplicated to automate MTs due to the so-called "add-on" design, resulting in a low-cost AMT. Furthermore, the high efficiency of AMTs has to be mentioned, which is higher than that of conventional ATs.

In addition to the add-on approach, there are AMTs that were designed as such. One example is the multigroup transmission of the Smart (Figure 23.24), which could not be shifted manually. The automation allows for a very small transmission design.

23.5.3.3 Continuously Variable Transmission

The transmission ratios of CVTs can be adjusted freely within design limits. The CVTs have an infinite number of gears. This is advantageous in combination with the characteristics of a combustion engine, since operation in an optimal point depending on the driving situation is possible. A constant engine speed can also be achieved through continuous change in transmission ratio if the speed is changed.

The so-called variator is the core of the CVT. It is used to change the transmission ratio between initial and final ratio continuously. The most common mechanical variators include a rotating element at input as well as output that are fixed through a linking element. The contact radii between linking element and input or output element can be varied to adjust the ratio.

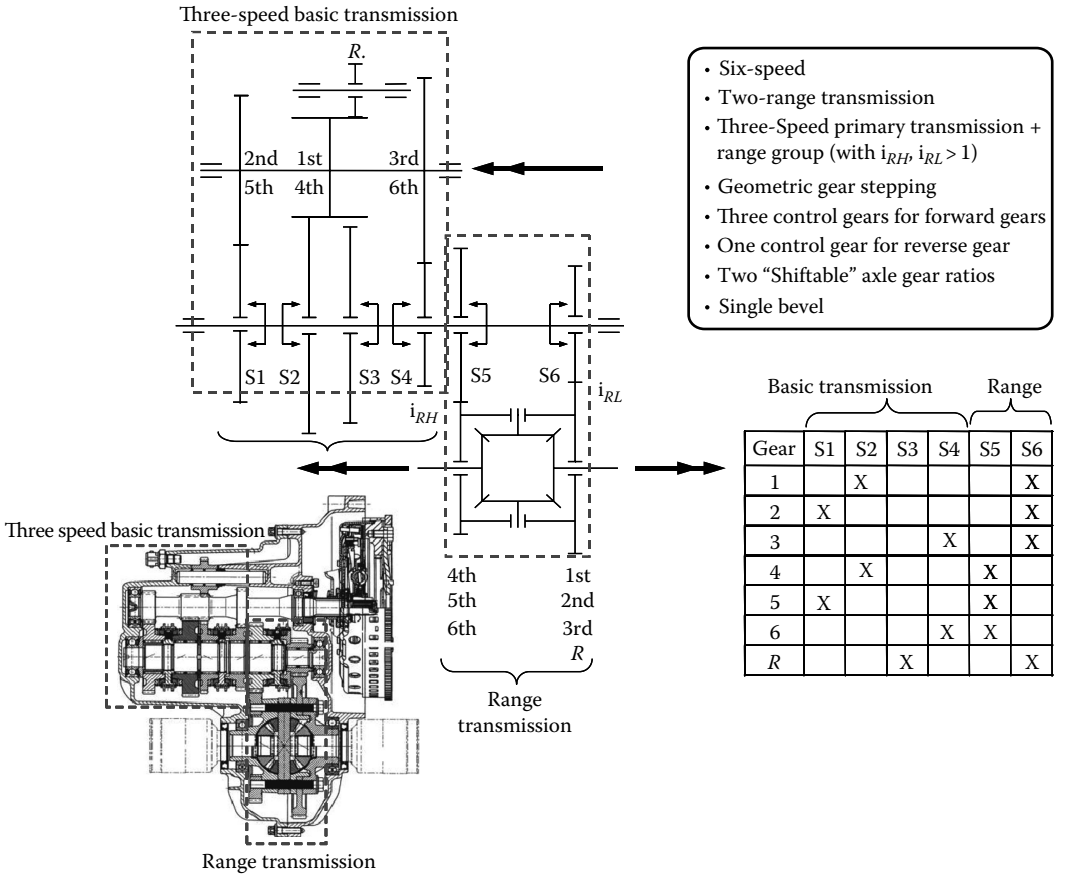
Most of the standard CVTs are so-called belt drive transmissions (Figure 23.25). A chain or belt is used as a linking element between two conical disks at the input and two at the output. Adjusting the axial distance of the pairs of disks changes the rotation radii of the belt or chain, which define the ratio.

The advantages of free ratio selection, however, are often compensated by a comparatively high energy input to generate the clamping force of the variator despite having made progress. Furthermore, these designs have not been very successful at least in European markets due to the unusual engine behavior of vehicles with CVTs.

Extended details are found in Refs. [32,33].

23.5.3.4 Double-Clutch Transmission

DCTs are designed as spur gear transmissions, similar to MTs. The difference is that there are two clutches at the transmission input. The transmission is divided into two partial transmissions: one for the odd (1, 3, 5,…) and one for the even gears (2, 4, 6,…). The partial transmissions are



	Basic transmission				Range	
Gear	S1	S2	S3	S4	S5	S6
1		X				X
2	X					X
3				X		X
4		X			X	
5	X				X	
6				X	X	
R			X			X

FIGURE 23.24 Layout of the AMT by Getrag, used in the Smart. It is a multigroup transmission, that is, two transmissions engage one after the other. Both transmissions are engaged at the same time when shifting from 3 to 4 or from 4 to 3, which would not be possible manually.

connected to the output. One clutch each is responsible for one partial transmission; the clutch for the odd gears as starting clutch is usually larger in diameter. During operation, one clutch is closed, while the other does not transmit torque. To reduce the influence of the rotational irregularities of the combustion engine and the load-alteration effect, the clutch is usually operated at a very low slip.

The advantage of this transmission design compared to MT or AMT designs is the fact that gears can be changed without interruption of traction because the torque is transferred from one clutch to the other during shifting (power shifting or clutch-to-clutch shifting). The following gear is preselected in the unloaded partial transmission, so the shifting times are short. The clutches are usually controlled hydraulically by a complex mechatronic control unit. DCTs for higher torques are equipped with wet multidisk clutches with oil cooling, whereas smaller DCT designs aiming for efficiency have dry clutches. Figure 23.26 shows the schematic diagram of a seven-speed DCT for front transverse installations with dry clutches.

A more detailed example is shown in Section 23.6.3.4. Extended details are found in Refs. [34–37].

23.5.3.5 Automatic Transmission

ATs are usually based on planetary transmissions with various combinations of gearsets and a hydrodynamic torque converter as starting device. A simple planetary transmission consists of a

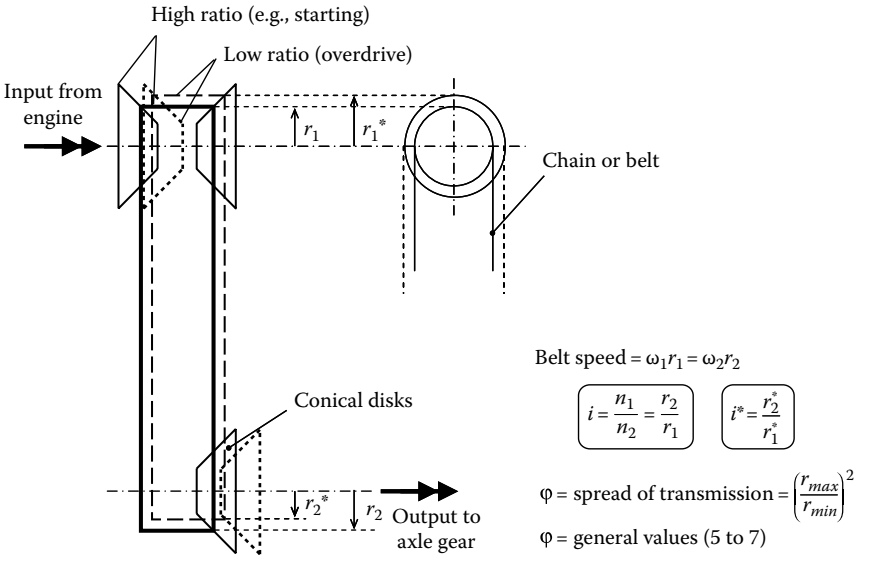


FIGURE 23.25 Function and ratio adjustment by means of a belt drive transmission. Through specific axial displacement of the conical disks, the rotation radii of the belt or chain at input and output are varied, resulting in a change in the ratio of the variator.

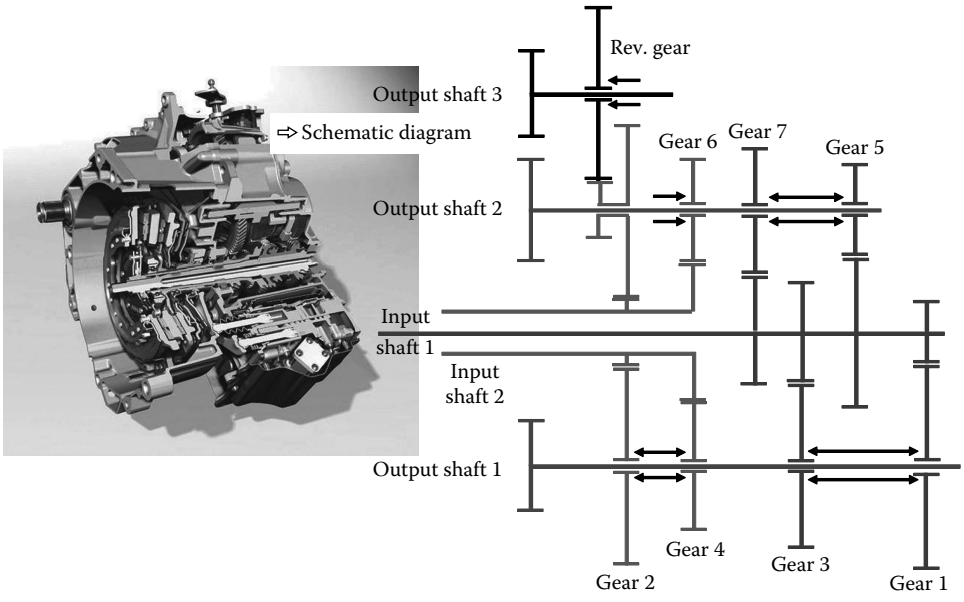


FIGURE 23.26 Layout of the seven-speed DCT DQ200 for front transverse installations by Volkswagen with dry double clutches. Both input shafts have one clutch and are responsible for either odd or even gears. The power flows from the respective input shaft to one of the output shafts and on to the axle gear with integrated differential. The synchronization units (double-headed arrows) are arranged in such a way that more than one gear step lies between the pairs of gears that are controlled by one synchronization unit. Therefore, gears can always be preselected for a gear change of one gear step (up- and downshifts).

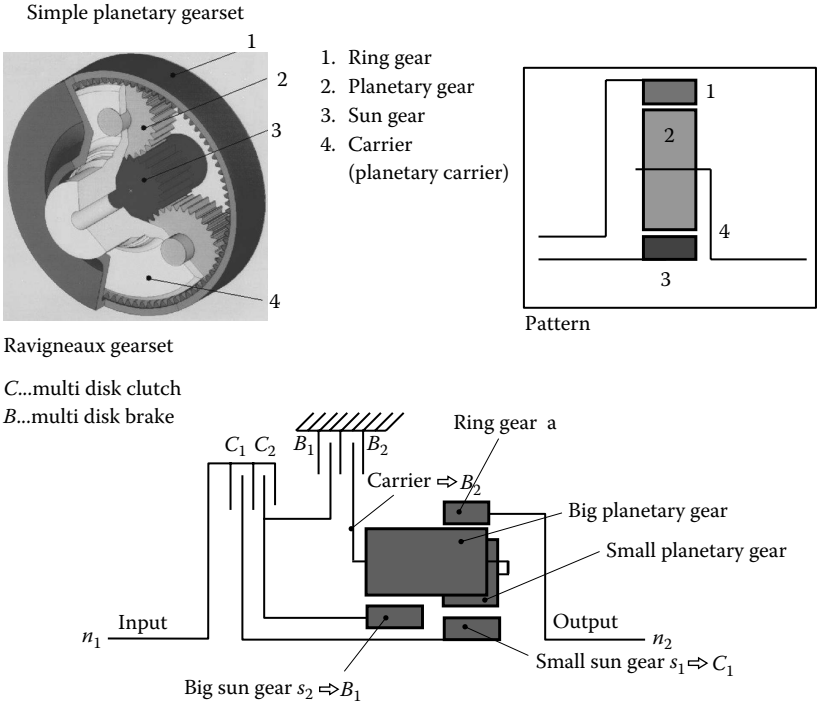


FIGURE 23.27 Structure of a simple planetary gearset and a Ravigneaux gearset. The Ravigneaux gearset consists of a simple planetary gearset and a second gearset with the planets meshing the planetary gears instead of the ring gear of the first gearset. The carriers of both gearsets are connected to each other.

central sun gear, an exterior ring gear, and several planetary gears between the sun and ring gears. The axes of the planetary gears are connected by a carrier (Figure 23.27). Sun gear, ring gear, and carrier are each connected to a shaft through which power can be transmitted or with which the respective component can be held to the transmission casing. All shafts are arranged coaxially. Different ratios can be achieved depending on the respective input and output shaft or which shaft is fixed. Most of the ratios, however, are not suitable for use in vehicles. For this reason, different gearsets were developed that allow for better transmission ratios. This, for instance, includes the Ravigneaux gearset represented in Figure 23.27.

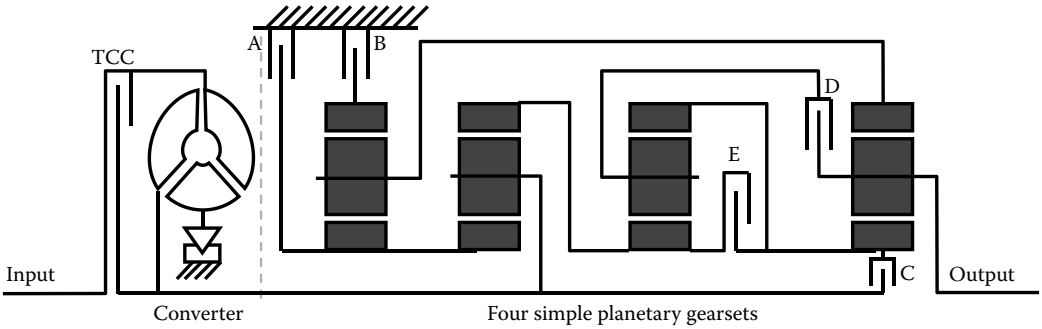
There are usually several gearsets in modern ATs. Their shafts can be connected by clutches or held by brakes. Depending on which of these so-called shift elements is activated (i.e., transmits torque), different ratios are possible. In this way, up to eight forward gears are available in modern transmissions. The power flow is partially split in a complicated way, whereby the gear system can be used optimally and high power density can be obtained. Figure 23.28 exemplifies the layout of an eight-speed AT.

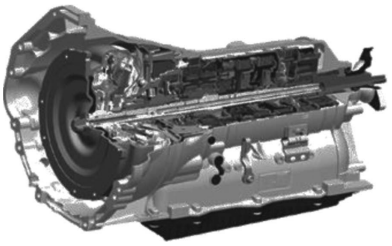
Wet multidisk clutches and brakes with oil cooling are installed as shift elements to change gears, like in DCTs as overlapping shifting. Thus, ATs also allow shifting without interruption of traction force. Shifting is controlled mechatronically, again like in DCTs.

Extended details are found in Refs. [38–40].

23.5.4 Drive Shafts

Most automobile applications comprise a longitudinal shaft to deliver power from the engine or transmission to the other end of the vehicle before it is transmitted to the wheels. A pair of short drive shafts commonly is used to transfer the power from a differential, the transmission, or the transaxle to the wheels. All drive shafts are important elasticities in the driveline and are therefore





Eight forward gears + reverse gear due to five shift elements

• Closed shift element

Gear	Brakes		Clutches			Ratio	Gear Step
	A	B	C	D	E		
1	•	•	•			4.70	1.50
2	•	•			•	3.13	1.49
3		•	•		•	2.10	1.26
4		•		•	•	1.67	1.30
5		•	•	•		1.29	1.29
6			•	•	•	1.00	1.19
7	•		•	•		0.84	1.25
8	•			•	•	0.67	Spread
R	•	•		•			7.05

FIGURE 23.28 Layout of the eight-speed stepped AT by the company ZF Friedrichshafen consisting of four simple planetary gearsets and five shift elements A–E. According to the desired gear, three of the shift elements are closed by hydraulically actuated pistons, while the other two are opened. A hydrodynamic converter with TCC is used as starting device.

responsible for the natural dynamic behavior concerning torsional vibrations. They are also sensitive to NVH effects and influence the "dynamic and acoustical quality" of the total system [41].

23.5.4.1 Rear-Wheel Drive and Front-Wheel Drive

Inside a front-engined, rear-wheel drive (RWD) vehicle, a long drive shaft is required to transmit the power to the wheels. Two applications dominate

- The torque tube with a single universal joint
- The Hotchkiss drive with two or more joints

Most of these vehicles have a clutch and gearbox (or transmission) mounted directly onto the engine's output shaft with a drive shaft leading into a final drive inside the rear axle.

The drive shaft connecting the gearbox to a rear differential is called a propeller shaft, or propshaft. A propshaft assembly contains of the propeller shaft, a slip joint, and one or more universal joints. Where the engine and axles are separated from each other, as on a four-wheel drive (4WD) and inside RWD vehicles, it is the propeller shaft that transmits the drive force from the engine to the axles.

A drive shaft connecting a rear differential to a rear wheel may be called a half shaft. The name derives from the fact that two such shafts are required to form one rear axle.

Several different types of drive shafts are used in the automotive industry:

- One-piece drive shaft
- Two-piece drive shaft
- Slip-in-tube drive shaft

Further details are found in Ref. [42].

23.5.4.2 Four-Wheel Drive and All-Wheel Drive

These two applications evolved from the front-engine RWD layout. A new type of transmission called the "transfer box" is placed between the transmission and the final drives inside both axles. This action does split the drive into the two axles and may also include reduction gears, a dog clutch, or differential. At least two drive shafts are used, one from the transfer box to each axle. In some larger vehicles, the transfer box is centrally mounted and driven directly by a short drive shaft.

Modern light cars with all-wheel drive (AWD) may use a system that more closely matches a front-wheel drive (FWD) layout. The transmission and final drive for the front axle are combined inside one housing mounted at the side of the engine, and a single drive shaft connects the rear axle.

A more detailed example is shown in Section 23.6.1 (Figure 23.31) and in Ref. [43].

23.5.4.3 Cardan (Universal) Shaft

In case of drivelines with power transmission that is parallel shifted or has an angular offset, Cardan (universal) joints are usually used. Especially inside the automotive industry, Cardan joints are common, since the spatial location of the driven and non driven end is variable due to length compensation of the Cardan shaft. Nevertheless, it must be regarded that Cardan joints can also be the source of excitations, which can lead to resonance-like amplifications (load peaks) inside the driveline. This excitation is due to the nonuniform behavior that can be described by a π-periodic (mostly sinusoidal shape) change of the ratio.

The equations for the calculation of the momentums of universal shafts can be taken from the technical literature [44].

Usually, two "ratios" (meaning the transformation of the degree of imbalance between two joints and the interposed shaft), two spring–damper elements, and one mass moment of inertia are considered, and some principal parameters are required for the calculation of the excitation mechanism (apart from the usual inertias, stiffnesses, and dampings) as shown in Figure 23.29.

23.5.5 Differential Gear

A differential is a device, usually but not necessarily consisting of gears, capable of transmitting torque and rotation through three shafts, almost always used in one of the two possible directions (ways). In one way it receives one input and provides two outputs (this is found in most automotive applications). In the other way it combines two inputs to create one output, which is the sum, difference, or mean of the two inputs.

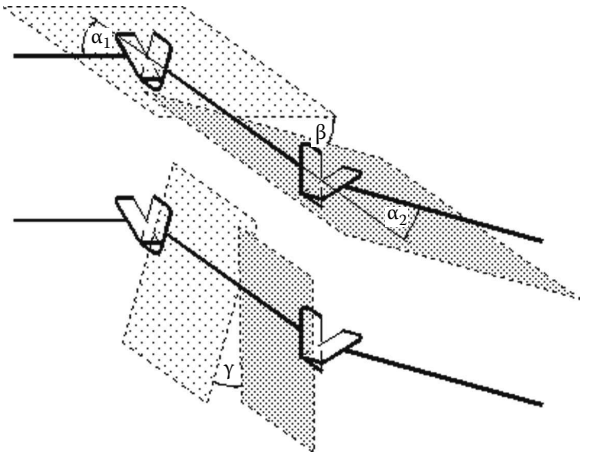


FIGURE 23.29 Angle definition inside the Cardan shaft, where α_1, α_2 = tilt angles at both connectors; β and γ = level angles (angular difference between two fork levels).

Inside automobiles and other wheeled vehicles, the differential allows each of the driving road wheels to rotate at different speeds, while for most vehicles supplying equal torque to each of them.

The purpose of the differential can briefly be described as follows. The wheels of a vehicle rotate at different speeds, especially when turning corners. The differential is designed to drive a pair of wheels with equal force while allowing them to rotate at different speeds. Inside vehicles without a differential, such as karts, both driving wheels are forced to rotate at the same speed, usually on a common axle driven by a simple chain-drive mechanism. When cornering, the inner wheel needs to travel a shorter distance than the outer wheel, so with no differential, the result is the inner wheel spinning and/or the outer wheel dragging, and this leads to a difficult and unpredictable handling, damage of tires, and possible failures of the complete drivetrain.

The superimposition of lateral and torsional vibrations sometimes occurs in differential gears in connection with the dynamic behaviors of the adjacent input and output shaft sections. Thus, such subsystems come up with typical dynamic effects that can only be solved in superimposed vibration models [45].

Extended details of modern differential gear concepts are found in Refs. [46,47].

23.6 SIMULATION OF THE DYNAMIC DRIVELINE BEHAVIOR: EXAMPLES

A quantitative statement concerning the current driveline excitation and system response (also considering nonlinear and transient effects) can only be made by the help of a "simulation in the time domain." A "steady-state simulation" (simulation in the frequency domain) is helpful in case of linear systems excited by harmonic excitation torques only; in this case the maximum amplitudes of the response functions (like torques) are computed for every speed step. As most of the effects in a vehicle driveline are nonlinear and also time transient (e.g., gear engagements), the applied simulation method is based on time-domain methods [9,10].

23.6.1 Drivelines: Studying Major Dynamic Effects in a Vehicle

Before starting the analysis of dynamics and NVH effects in drivelines in more details, let us study the sources of noise and vibrations that occur in a driveline.

Usually a driveline, excited by the typical fluctuations due to the reciprocating engine, consists of the flywheel (single- or twin-mass flywheel), the clutch, the manual or AT including all shafts, the differential gear with the side shafts, wheels, and tires. There are different drive concepts like

- FWD
- RWD
- AWD or 4WD

Figure 23.30 shows the typical drive elements of a RWD passenger car powertrain with the corresponding NVH phenomena that might occur [17,48].

If we consider truck powertrains (4WD or AWD), we observe similar NVH effects as presented in Figure 23.31 [48].

The numbers in Figure 23.31 refer to typical NVH effects (NVH phenomena) as presented in Table 23.2 (including the characteristic frequencies that are usually observed due to the mentioned effects). In the table, further references are mentioned referring to a more detailed description of NVH effects.

It is a must for the R&D engineers to design the vehicle driveline in such a way that the dynamic effects are analyzed a priori before prototypes are built or the first vehicles are tested in detail. Thus, it is crucial to be aware of the characteristics of the powertrain and driveline concepts concerning the state-of-the-art vehicles but also with reference to future vehicle design. The transmission concept is one of the key concepts of a vehicle powertrain. Usually, the transmission

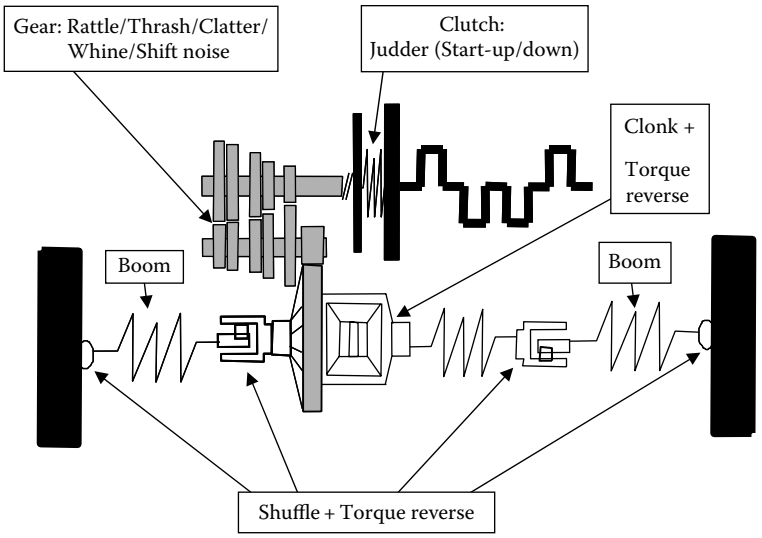


FIGURE 23.30 Typical NVH phenomena in a passenger car powertrain (FWD).

design has to be combined with the development of clutch systems or multi-mass flywheel systems including the damper design.

Furthermore, MTs lose importance in the global market. Typically, there are NVH problems in driveline configurations with manual MTs. Therefore, the development and NVH optimization of ATs become more and more significant. The customer attributes related to existing design concepts of various ATs are visualized in Figure 23.32 [57]. Obviously, no concept really fulfills all attributes in an optimum way. The customer attributes tend to a positive rating when considering a more modern design concept like higher numbers of gears in AT, or new concepts like DCT and CVT (cf. Section 23.5.3). However, concerning the dynamics, we have to reckon with new effects like self-excitations in the twin clutch due to negative friction gradients depending on the slip in case of a DCT or rubber band effects in case of a CVT. That is why modern transmission concepts require a very intense a priori dynamic analysis to find the best NVH rating. Apart from the dynamic effects, the driveline concept always has to meet all modern fuel economy requirements depending on the different classes of vehicles including best performance and efficiency. In most of the cases, the R&D engineer has to cope with typical target conflicts—and it is up to him to solve these system problems using modern computer simulation technologies, high-quality models, and testing results [58].

23.6.2 Nonlinear Behavior of Drivelines

Drivelines imply the interconnection of system elements with nonlinear effects as found in the clutch or two-mass flywheel (DMF = dual/two-mass flywheel), or during the synchronizing process, or as backlash in the total drive system (gear stages, transmission, and shafts), or as slip-dependent wheel–tire–road contact. Some typical characteristics are summarized in Figure 23.33. That is why dynamic effects dedicated to specific drive elements are often only understandable when the overall system behavior of all connected drive elements is considered.

23.6.2.1 Introducing a Simplified Calculation to Evaluate Gear Rattling

Gear rattling is a nonlinear effect that can considerably impair the inner room noise comfort of vehicles. For the optimization of this "NVH" phenomenon (cf. Section 23.6.1), the whole driveline must be regarded, because gear rattling is excited by external forces due to the engine torque fluctuations, while the rattle noise occurs inside of the gearbox due to the backlash in unloaded gear stages that are

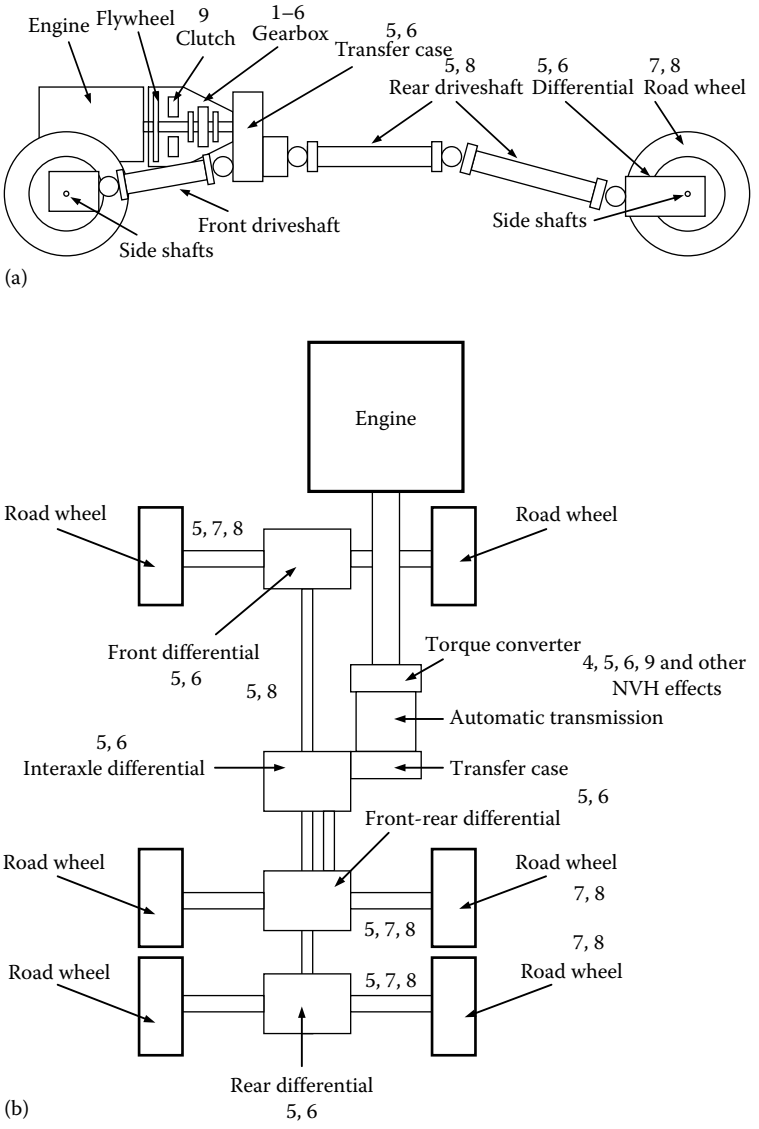


FIGURE 23.31 Typical NVH phenomena in a truck powertrain (4WD/AWD). (a) With an MT (4 × 4 drive) and (b) with an AT (6 × 6 drive).

not in the torque path of the transmission train. This means that a "gear problem" is not a localized problem of the gear but becomes a problem of the total drive system configuration. As a degree for the audible rattle noises, it is recommended to use minimized models first before starting with comprehensive simulation studies. Such models may be evaluated by the introduction of a so-called rattling factor [59], which allows the comparison of various gear design concepts in a very early R&D step.

Rattle vibrations originate inside toothing areas and are transported via gear shafts and bearings to the bearing housing and further over the vehicle body to the chassis. In order to compare the theoretical model with the acoustic measurements, a simplified noise model is used assuming that the sound power due to rattling is proportional to the power loss, which occurs due to shocks inside the toothing area [59]. If the loss energy is added over a time period and applies it to this time period, you receive the lost energy. The most significant influence on rattling noises is presented in Figure 23.34. It is obvious that the external excitation has the major influence compared to the backlash within the gear stages.

TABLE 23.2
NVH Effects and the Corresponding Frequency Range

Effect No.	NVH Phenomena	Typical Frequency Bandwidth
1	***"Rattling"/"Thrashing"*** [49–51]	Excitation **~20–120 Hz**
	During acceleration or coasting	Noise **~500–3000 Hz**
2	***"Clattering"/"Chattering"*** [52]	**~8–20 Hz**
	e.g., in idle position	Also higher frequencies due to noise
3	***"Clashing"***	**~25–80 Hz, 300–600 Hz, 1000–2000 Hz**
4	***"Gear Shift Noise"***	**~25–80 Hz, 300–600 Hz, 1000–2000 Hz**
5	***"Clonk"/"Clunk"*** [53,54]	**~300–600 Hz, 1000–2000 Hz**
	"Torque Reverse"/"Load Change"	**~1–10 Hz, 50–150 Hz, 1000–2000 Hz**
6	***"Gear Whining"*** [55]	**~500–4000 Hz**
7	***"Shuffle"/"Tip-in Back-out"***	**~1–10 Hz**
	Also known as "Bonanza" effect	
8	***"Boom"***	**~20–50 Hz, 100–250 Hz**
9	***"Judder"/"Shudder"*** [56]	**~10–90 Hz**
	Slip + friction effects depending on the coupling design	

The effect numbers refer to Figure 23.31.

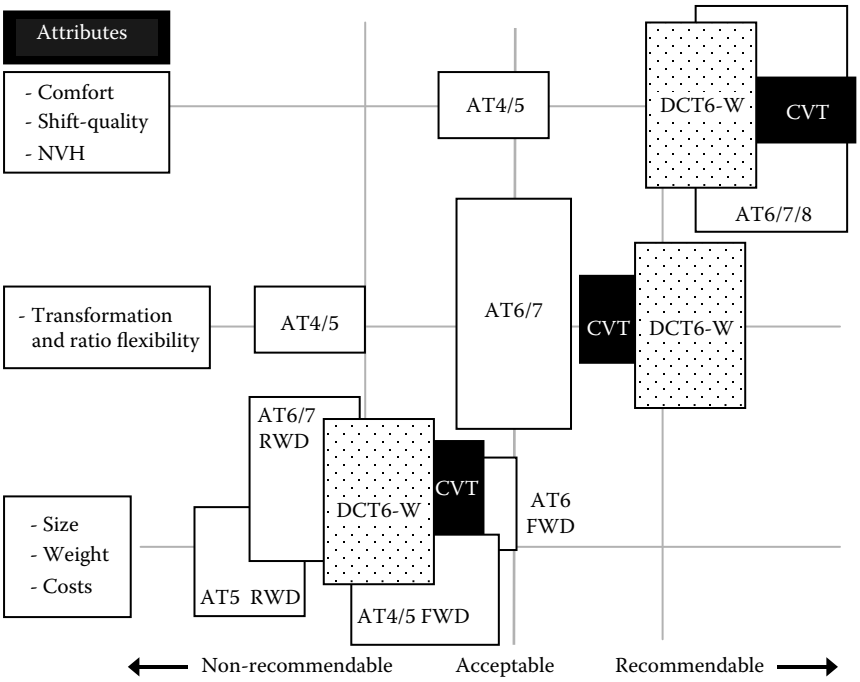


FIGURE 23.32 Customer attributes concerning existing AT concepts. (AT with a torque converter, DCT [W = wet clutch], CVT, and 4/5/6/7/8 = number of gears).

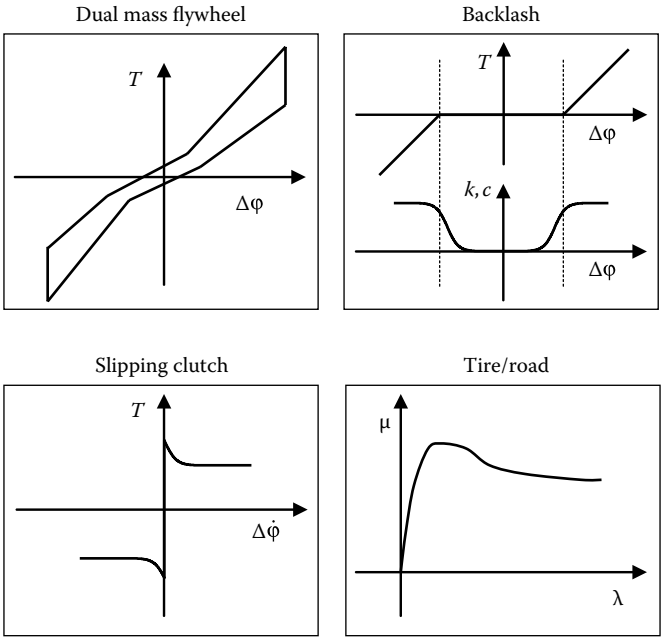


FIGURE 23.33 Examples of nonlinear effects inside a driveline.

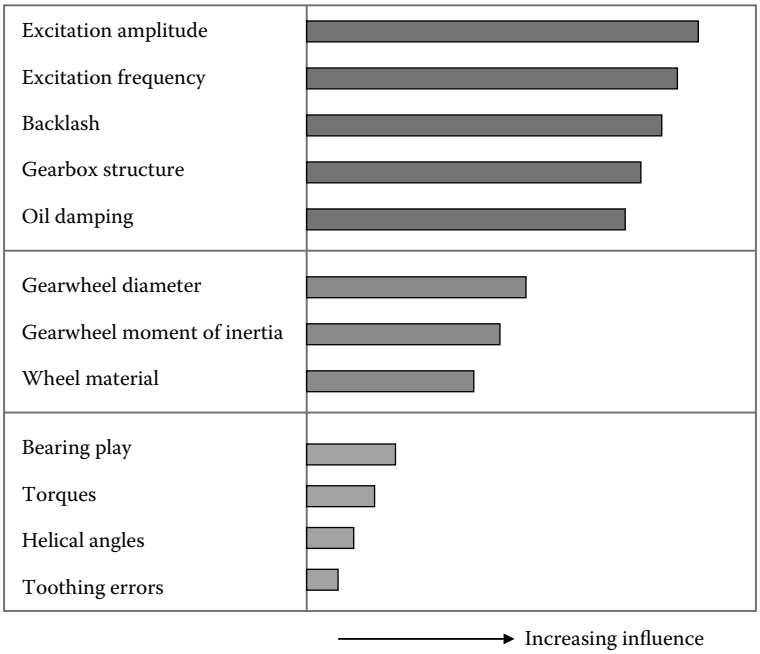


FIGURE 23.34 Influence of rattling effects depending on typical parameters. (From Küçükay, F., Berechnung und Optimierung der Rasselschwingungen bei Schaltgetrieben, in: *Berechnung im Automobilbau*, VDI Report No. 699, Düsseldorf, Germany, pp. 593–630, 1988.)

The calculation of the rattling factor R is based on the summation of the impact power over all p loose gear wheels in the total transmission. First, it is recommendable to normalize the input speeds n_{ej} of all loose "driving" gear wheels with the radius r_e (index = j) to the input speed n_{e1}. Taking r for the radius and J for the inertia of the "driven" loose gear wheel, we get the rattling number R_z as follows:

$$i_j = \frac{n_{e1}}{n_{ej}} \tag{23.12}$$

$$R_z = \sum_{j=1}^{p} J_j \frac{r_{ej}}{r_j^2} \frac{1}{i_j} \tag{23.13}$$

With the help of R_z, gearboxes with different structures on the one hand but same attributes (like backlash, transmission oil) on the other hand can be compared according to their principal rattle sensitivity. For the comparison, it is useful to choose a reference gearbox, whose rattling number in idle state R_{z0} is used as reference number. If the rattling number R_z is now normalized with R_{z0}, we get the dimensionless rattling factor R:

$$R = \frac{\sum_{j=1}^{p} J_j \, (r_{ej}/r_j^2)(1/i_j)}{R_{z0}} \tag{23.14}$$

Further details of this method are described in Ref. [59].

23.6.2.2 Simulation Models to Analyze the Gear Rattle Behavior

Further calculation steps supported by advanced simulation models are required, if the total driveline is optimized concerning the gear rattle effects but also with reference to other NVH effects that may lead to target conflicts when introducing any improvements or innovations. In order to analyze most of the NVH problems, a reduced simulation model can be used as shown in Figure 23.35 before creating comprehensive models [17]. See Section 23.4.1 for more explanations referring to the understanding of the energy presentations of vibration modes.

Based on the reduced model, the important gear branches with the loose gear wheels are integrated in the simulation model (Figure 23.36) in order to calculate some typical NVH rating functions [17,48] such as the following:

- RMS values (RMS = root mean square)of the loose gear wheel accelerations and fluctuations
- Number of impacts and classifying the time signals to identify NVH patterns
- RMS values of the impact energy, also the summation of all impact energies for every gear stage
- Superimposed analysis of the total system also considering statistical parameters and distributions standing for the variation of backlash but also for the clutch/DMF characteristics, further elasticities in the driveline, etc.

Statistical methods [17] are useful to consider

- Data variations
- Data uncertainties
- Data sensitivities
- Tolerances due to the manufacturing process
- Parameter range that is acceptable (and realizable) and can be taken for any further improvements

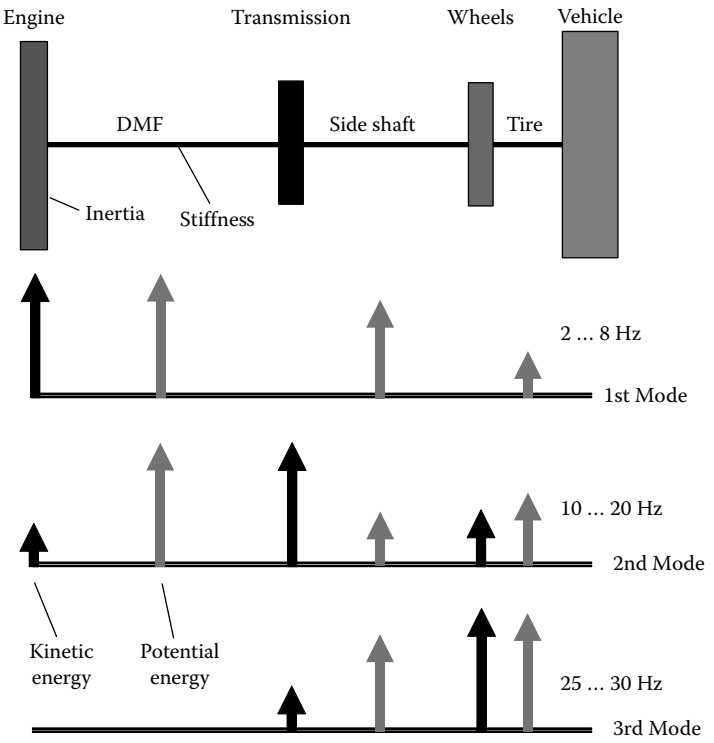


FIGURE 23.35 Reduced torsional vibration model of a driveline to discuss the basics of NVH problems. (From Laschet, A. and Fan, J., Ermittlung der NVH-Sensitivität durch geeignete CAE-Modellierung, in: S. Pischinger and H. Wallentowitz (eds.), *12. Aachener Kolloquium Fahrzeug- und Motorentechnik*, Technical University, Aachen, Germany, pp. 719–733, 2003.)

Therefore, the rating of NVH phenomena should also lead to a statistically based rating function [17]. The input values are no longer fixed parameters but parameter distributions (e.g., Gaussian, rectangular distributions). The sharper the rating distribution, the better the quality repeatability.

23.6.2.3 Advanced Evaluation of Superimposed Nonlinear Characteristics

As NVH effects cannot be rated just by taking single phenomena independently of other effects, it is crucial to create a rating matrix with the most significant characteristics including some aspects on efficiency and costs (Figure 23.37) [60]. Therefore, a more-dimensional rating system is required to investigate all typical target conflicts in order to come to an optimized solution. For this kind of computation studies, statistical methods are mandatory.

23.6.3 Driveline Modeling and Simulation of the Shifting Behavior

23.6.3.1 Targets and Benefits of Shifting Simulation

In manual gearboxes, each gear stage contains at least one pair of gears with different transmission ratios. Changing gears and implies the engagement disengagement of dog clutches, which positively connect idler gears to the gearbox shafts. The same principle takes place in many automated gearbox types (e.g., in DCTs). The required speed synchronization before engaging the dogs is achieved either by mechanical synchronizers (MT, DCT) or by external means such as the engine or gearbox brakes in heavy-duty vehicle powertrains.

The opening and closing of the dog clutches is the most important part in the shifting process. The quality of these operations has significant influence on the shift quality and comfort. Some aspects are of importance as presented in Table 23.3.

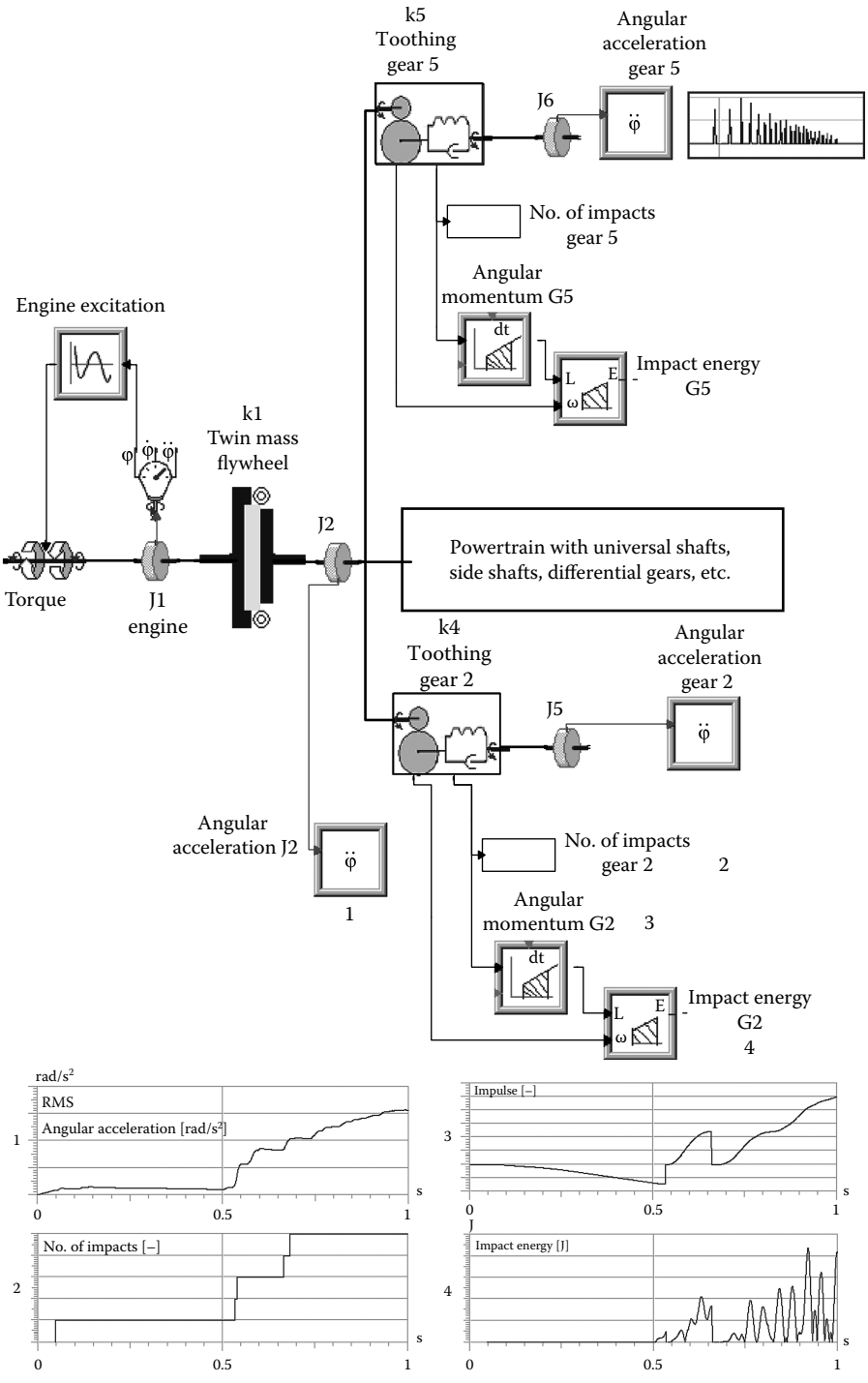


FIGURE 23.36 Simulation model realized by ITI®-SIM 3 to create typical NVH rating functions.

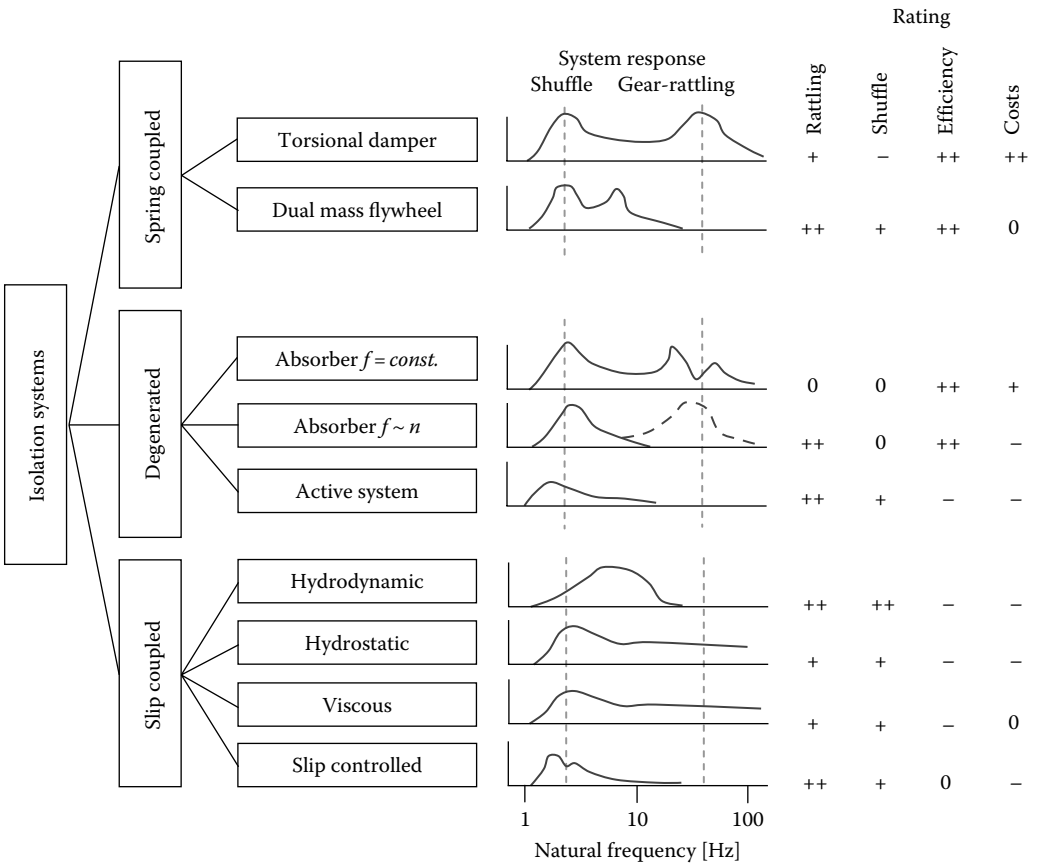


FIGURE 23.37 Rating of vibration damping and isolation in order to improve more than one NVH phenomenon in drivelines. (From Reik, W., Torsionsschwingungsisolation im Antriebsstrang—ein Wertungsversuch, in: *Torsionsschwingungen im Antriebsstrang*, Baden-Baden, LuK, Bühl, Germany, pp. 119–138, 1990.)

TABLE 23.3
Opening and Closing of Dog Clutches: Criteria and Influences

Criterion	Influence On
Speed equality during dog engagement	Clashing, vibration excitation on the powertrain, noise excitation, component wear, comfort perception on the shifting lever
Response forces during dog clutch engagement	Comfort perception by the driver; smoothness and efforts of shifting
Load torques during dog clutch disengagement	Excitation of vibrations on the driveline; disturbance of the subsequent shift process

23.6.3.2 Development of Shifting Simulation Methods

Modeling a powertrain for shifting simulation purposes is a step-by-step procedure, which starts at the innermost component in the shift scenario—the dog clutches and synchronizers. The resulting models are validated against test data acquired from measurements on component test rigs. This ensures the proper behavior of the component and allows tuning of the limited number of not or hardly measurable parameters such as damping or friction characteristics.

In a next step, the component model is included into a larger-scale model, such as a gearbox model, for which the procedure of validation against test data is carried out in the same manner. Finally, a powertrain model is the result of the last stage of this process.

23.6.3.3 Example: Manually Shifted Powertrain

The following example considers the shifting in a manual gearbox. The aim of this simulation task [61] is the optimization of the shift through, in particular for the lower gears. Potential comfort problems such as scratching under low-temperature conditions are of particular interest as well.

The simulation task follows a three-stage outline according to the approach as described earlier:

- Synchronizer modeling and validation
- Gearbox modeling and validation
- Complete powertrain modeling and validation

The final stage of the model development leads to a reproduction of the gearshift behavior as observed on the real vehicle. Once validated, on this model base there is a high certainty that a change of system parameters in the model will affect the simulated shifting behavior in the same way as the shifting behavior observed on the real vehicle.

Figure 23.38 shows a particularly bad shift from the perspective of the moving sliding sleeve. The effect seen here is clashing, which appears for very low oil temperatures and small shift forces. The increased drag torque leads to the formation of a speed difference during the short time interval between the locking teeth giving way and the sleeve reaching the selector teeth. Due to the speed difference, the sleeve is pushed back several times before the teeth engage, which is felt as a typical NVH effect by the driver (so-called scratching or clashing).

23.6.3.4 Example: Dual-Clutch Transmission

Dual-clutch concepts integrated into the driveline design combine a synchronized gearbox with automated shifting and provide the additional benefit of uninterrupted traction, resulting in a symbiosis of efficiency, comfort, and performance.

DCTs have gearboxes with two input shafts, each having its own clutch, two separate countershafts being engaged for even or odd gears, respectively, and one common output shaft (Figure 23.39). When shifting, a gear on one of the countershafts is engaged and the respective clutch is closed. On the other countershaft (with the corresponding clutch open), the next gear can be engaged during normal drive. Once the dog clutch for the new gear became engaged, the torque flow is handed over from one countershaft to the other by opening one and closing the other clutch simultaneously. The comfort of the shift is mainly determined by the way how smoothly the transition between the two gears takes place. General descriptions are also given in Section 23.5.3.4.

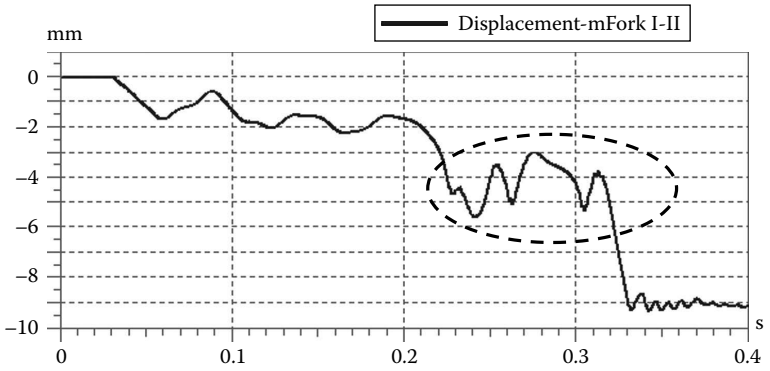


FIGURE 23.38 Clashing in a cold gearbox with low shift forces.

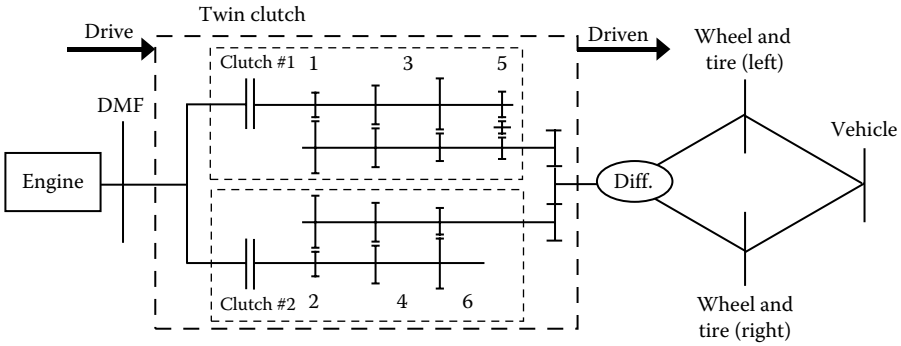


FIGURE 23.39 Example of a FWD driveline with dual-clutch transmission (=DCT).

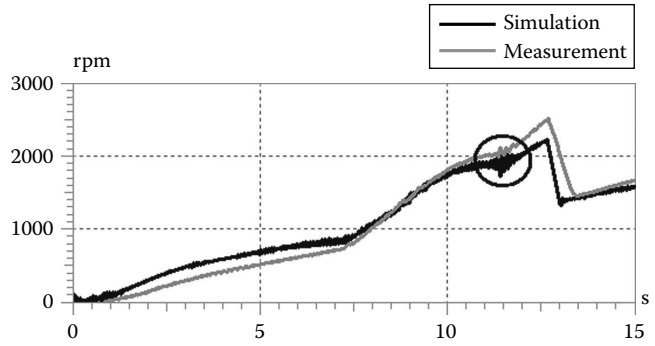


FIGURE 23.40 DCT gearbox input shaft speeds (simulation and measurement).

The example models and results are discussed in more details in Refs. [62,63].

A major analysis issue for shifting comfort in DCTs is the transition between the two branches in the transmission when shifting from one gear to the other. The simulation model is used in order to test the clutch actuation off-line and to evaluate the resulting quality of the gearshift at an early phase of the development cycle.

Figure 23.40 shows a comparison of measured and simulated input shaft speeds of the gearbox. The gear change takes place in the time interval from 10.5 to 12 s. The presented speeds correspond to the input shaft, which is disengaged during the shift. The focus of this analysis is principally on the gearbox vibrations excited during the gear change (marked by a black circle in the figure).

23.7 SUMMARY

The dynamic behavior of drivelines becomes critical when regarding stress, loads, excitations, and above all effects due to NVH. Such dynamic criteria stand for the level of quality and acceptance from the customer side. Therefore, it is crucial to compute the most important and critical dynamic effects in advance using simplified models on the one hand but also more comprehensive simulation methods and software tools on the other hand. Particularly, a practice-oriented rating system is required to evaluate the concepts as early as possible during the R&D process with best computing quality.

In spite of the advanced application of modern simulation technologies, one important connecting link must not be missing: the *engineer* is always the center of the product development chain. Engineerlike thinking and acting remain significant and central essential parts of the complex assessment of technical systems.

ACKNOWLEDGMENTS

The authors wish to thank Mr. Andreas Stöcker from ARLA Maschinentechnik GmbH, Wipperfürth, Germany for his assistance doing the calculations, the detailed analyses, and preparing the graphical presentations. The authors particularly thank Dr. Tobias Kassel from the Institute of Automotive Engineering (Braunschweig University of Technology, Germany) for his hard work, numerous ideas and suggestions in the development of this chapter and in addition also Mr. Gerrit Knabe and Mr. Jan Häfele for numerous technical suggestions and editorial notes. Furthermore the authors like to thank Mr. Andreas Abel from ITI GmbH, Dresden, Germany for his contribution for Section 23.6.3.

REFERENCES

1. Laschet, A. (2009). The overall system behaviour of vehicle powertrains. In: A. Laschet et al.: *Systemanalyse in der Kfz-Antriebstechnik V* (pp. 5–8). Renningen, Germany: Expert Verlag.
2. De Silva, C.W. (2000). *Vibration: Fundamentals and Practice*. Boca Raton, FL: CRC Press.
3. Piersol, A.G. and Paez, Th.L. (2009). *Harris' Shock and Vibration Handbook*, 6th edn. New York: McGraw-Hill.
4. Dresig, H. and Holzweißig, F. (2009). *Maschinendynamik*, 9th edn. Berlin, Germany: Springer-Verlag.
5. Dresig, H. (2006). *Schwingungen mechanischer Antriebssysteme*, 2nd edn. Berlin, Germany: Springer-Verlag.
6. Shigley, J.E., Mischke, Ch.R., and Budynas, R.G. (2004). *Mechanical Engineering Design*, 7th edn. New York: McGraw-Hill.
7. Lechner, G. and Naunheimer, H. (1999). *Automotive Transmissions*. Berlin, Germany: Springer Verlag.
8. Robert, B. (2007). *Automotive Handbook*, 7th edn. (Section "Drivetrain", pp. 756–773). Chichester, U.K.: John Wiley & Sons.
9. Laschet, A. (1988). *Simulation von Antriebssystemen. Fachberichte Simulation*, Vol. 9. Berlin, Germany: Springer-Verlag.
10. Laschet, A. (1995). Einsatz der Simulation im Rahmen antriebstechnischer Optimierungen. In: *Simulation in der Praxis*. Düsseldorf, Germany: VDI Report No. 1215, pp. 19–34.
11. Laschet, A. (2004). ARLA service for engineers—Torsional vibration analysis (TVA) of drive systems. Basic information on the model generation, analysis of excitability and interpretation of the dynamic behaviour of drivelines & powertrains, Report No. R-TVA/3.0. Wipperfürth, Germany: ARLA Maschinentechnik GmbH (www.arla.de).
12. Willumeit, H.-P. (1998). *Modelle und Modellierungsverfahren in der Fahrzeugdynamik*. Stuttgart, Germany: B.G. Teubner.
13. Niemann, G. and Winter, H. (1983). *Maschinenelemente*, 2nd edn., Vol. 2. Berlin, Germany: Springer-Verlag.
14. Smith, J.D. (1999). *Gear Noise and Vibration*. New York: Marcel Dekker, Inc.
15. Gregory, R.W., Harris, S.L., and Munro, R.G. (1963–1964). Dynamic behaviour of spur gears. *Proceedings of the Institution of Mechanical Engineers*, 178(Part 1): 207–226.
16. Laschet, A. and Berres, M. (2001). Strategien zur ingenieurmäßigen Modellgenerierung von Kfz-Antriebssträngen. In: A. Laschet et al.: *Systemanalyse in der Kfz-Antriebstechnik I* (pp. 3–24). Renningen, Germany: Expert Verlag.
17. Laschet, A. and Fan, J. (2003). Ermittlung der NVH-Sensitivität durch geeignete CAE-Modellierung. In: S. Pischinger and H. Wallentowitz (eds.): *12. Aachener Kolloquium Fahrzeug- und Motorentechnik*. Aachen, Germany: Technical University, pp. 719–733.
18. Rangwala, A.S. (2001). *Reciprocating Machinery Dynamics*. New York: Marcel Dekker, Inc.
19. Fayette, T.Ch. (1985). *Internal Combustion Engine in Theory and Practice*, Vol. 2. Cambridge, MA: MIT Press.
20. Heisler, H. (2002). Friction clutch. In: *Advanced Vehicle Technology* (pp. 37–59). London, U.K.: Butterworth-Heinemann.
21. Benatzky, C., Schober, M., and Schlager, G. (2009). A nonlinear transmission approach for dual mass flywheels. In: *Getriebe in Fahrzeugen 2009*. Düsseldorf, Germany: VDI Report No. 2071.
22. Zieglmaier, W. and Steinel, K. (2008). Clutch concepts for high engine torques in commercial vehicles. In: *CTI Transmission Symposium 2008*, Berlin, Germany. *Conference Book of 7th International CTI Symposium*.

23. Heisler, H. (2002). Hydrokinetic fluid couplings and torque converters. In: *Advanced Vehicle Technology* (pp. 98–116). London, U.K.: Butterworth-Heinemann.
24. Reik, W. (2007). Clutches, torque converters and dampers—Linking the engine and transmission in new drivetrain concepts. In: *CTI Transmission Symposium 2007*, Berlin, Germany. *Conference Book of 6th International CTI Symposium*.
25. Heisler, H. (2002). Semi- and fully automatic transmission. In: *Advanced Vehicle Technology* (pp. 117–192). London, U.K.: Butterworth-Heinemann.
26. Sutar, G.S. (2008). Modular shift system for five and six speed manual transmission. In: *CTI Transmission Symposium 2008*, Berlin, Germany. *Conference Book of 7th International CTI Symposium*.
27. Müller, P. and Helms, G. (2007). The new 6-speed manual transmission family for the MINI. In: *CTI Transmission Symposium 2007*, Berlin, Germany. *Conference Book of 6th International CTI Symposium*.
28. Ozaki, Y. (2006). The improvement technique on the performance of manual transmission. In: *CTI Transmission Symposium 2006*, Berlin, Germany. *Conference Book of 5th International CTI Symposium*.
29. Narita, R. (2009). Development of new 6-speed manual transmission for FWD passenger vehicles. In: *CTI Transmission Symposium 2009*, Berlin, Germany. *Conference Book of 8th International CTI Symposium*, Vol. 1, pp. 83–98.
30. Findlay, M. (2009). Next steps in automated manual transmission technology. In: *CTI Transmission Symposium 2009*, Berlin, Germany. *Conference Book of 8th International CTI Symposium*, Vol. 1, pp. 134–136.
31. Höhn, B.-R. (2009). Automated Manual transmission—The forgotten concept for the future? In: *Getriebe in Fahrzeugen 2009*. Düsseldorf, Germany: VDI Report No. 2071.
32. Schiberna, P. (2007). The newly developed second generation Audi multitronic. In: *CTI Transmission Symposium 2007*, Berlin, Germany. *Conference Book of 6th International CTI Symposium*.
33. Kurse, T., Meyer, C., and Poll, G. (2006). Influencing factors on the variator efficiency of belt-type CVT—Measurement data, simulation and design optimisation. In: *CTI Transmission Symposium 2006*, Berlin, Germany. *Conference Book of 5th International CTI Symposium*.
34. Tenberge, P. (2006). Dual-clutch transmission in planetary gearset design. In: *CTI Transmission Symposium 2006*, Berlin, Germany. *Conference Book of 5th International CTI Symposium*.
35. Gröhlich, H., Hadler, J., Schäfer, M., and Lindemann, K. (2009). DQ500—The new Volkswagen 7-speed dual clutch transmission for high torque. In: S. Gies and S. Pischinger (eds.): *18th Aachen Colloquium "Automobile and Engine Technology"*. Aachen, Germany: Technical University.
36. Bünder, C. (2009). "6DCT250"—The next generation high efficiency GETRAG PowerShift—Product for the world market. In: *CTI Transmission Symposium 2009*, Berlin, Germany. *Conference Book of 8th International CTI Symposium*.
37. Ebenhoch, M. (2009). 7-speed dual-clutch transmission with start-stop function and optional integrated all-wheel distributor system. In: *CTI Transmission Symposium 2009*, Berlin, Germany. *Conference Book of 8th International CTI Symposium*.
38. Usuki, K. and Uchida M. (2008). Development of new 7-speed automatic transmission for medium and large RWD vehicles. In: *CTI Transmission Symposium 2008*, Berlin, Germany. *Conference Book of 7th International CTI Symposium*.
39. Müller, J. and Leesch, M. (2008). New structure optimized 8-speed automatic transmission. In: *CTI Transmission Symposium 2008*, Berlin, Germany. *Conference Book of 7th International CTI Symposium*.
40. Tenberge, P. (2006). Dual-clutch transmission in planetary gearset design. In: *CTI Transmission Symposium 2006*, Berlin, Germany. *Conference Book of 5th International CTI Symposium*.
41. Reitz, A. (2003). Abstimmung der Schwingungsübertragung von Seitenwellen zur Verbesserung des Innengeräuschs in einem frontgetriebenen PKW. In: A. Laschet et al.: *Systemanalyse in der Kfz-Antriebstechnik II* (pp. 83–104). Renningen, Germany: Expert Verlag.
42. Nunney, M.J. (1998). Rear-wheel drive and front-wheel drive layouts. In: *Automotive Technology* (pp. 391–394). Warrendale, PA: Society of Automotive Engineers.
43. Nunney, M.J. (1998). Four-wheel-drive systems. In: *Automotive Technology* (pp. 427–437). Warrendale, PA: Society of Automotive Engineers.
44. Duditza, F. (1973). *Kardangelenkgetriebe und ihre Anwednungen*. Düsseldorf, Germany: VDI-Verlag.
45. Eulert, S., Klaffke, S., Lührs, G., and Martin, F. (2009). Experimentelle und numerische NVH-Analyse verzahnungsindizierter Biege- und Torsionsschwingungen am Beispiel einer Hinterachse. In: A. Laschet et al.: *Systemanalyse in der Kfz-Antriebstechnik V* (pp. 158–173). Renningen, Germany: Expert Verlag.
46. Heisler, H. (2002). Final drive transmission. In: *Advanced Vehicle Technology* (pp. 226–269). London, U.K.: Butterworth-Heinemann.

47. Pinschmidt, U. (2009). The new sport differential in the Audi S4, In: *Getriebe in Fahrzeugen 2009*. Düsseldorf, Germany: VDI Report No. 2071.
48. Laschet, A. (2004). CAE-Unterstützung bei der Analyse von NVH-Sensitivitäten in Getrieben und Nkw-Antriebssträngen. In: Ch. Erhard (ed.): *NVH bei modernen Nutzfahrzeugen*. Essen, Germany: Haus der Technik.
49. Rivin, E.I. (2000). Analysis and reduction of rattling in power transmission systems. SAE Technical Paper Series, No. 2000-01-0032.
50. Delprete, C. and Rosso, C. (2005). Numerical analysis of gear rattle. SAE Technical Paper Series, No. 2005-01-1786.
51. Kwon, O.J. and Lee, H.S. (2005). A study on the evaluation process of rattle noise considering the signal characteristics in frequency and time domain. SAE Technical Paper Series, No. 2005-01-2543.
52. Fujimoto, T. and Kizuka, T. (2005). Predictive calculation of idling rattle in manual transmissions. SAE Technical Paper Series, No. 2003-01-0678.
53. Oh, W. and Singh, R. (2005). Examination of clunk phenomena using a non-linear torsional model of a front wheel drive vehicle with manual transmission. SAE Technical Paper Series, No. 2005-01-2291.
54. Crowther, A.R., Zhang, N., and Singh, R. (2005). Development of a clunk simulation model for a rear wheel drive vehicle with automatic transmission. SAE Technical Paper Series, No. 2005-01-2292.
55. Cheng, Y., Abe, T., and Wilson, B.K. (2005). Automatic transmission gear whine simulation and test correlation. SAE Technical Paper Series, No. 2005-01-2290.
56. Mäki, R., Nyman, P., Olsson, R., and Ganemi, B. (2005). Measurement and characterization of anti-shudder properties in wet clutch applications. SAE Technical Paper Series, No. 2005-01-0878.
57. Petri, H., Najork, R., Eggert, U., and Krauss, Ch. (2004). Systeme, Konzepte, Lösungen—Automatikgetriebe 2004. In: *Getriebe in Fahrzeugen 2004*. Düsseldorf, Germany: VDI Report No. 1827, pp. 713–747.
58. Riel, A., Hasewend, W., Bogner, E., and Fischer, R. (2004). Modellierung von Fahrzeug und Antriebsstrang im gesamten Entwicklungsprozess. *ATZ*, 106(6): 522–531.
59. Küçükay, F. (1988). Berechnung und Optimierung der Rasselschwingungen bei Schaltgetrieben. In: *Berechnung im Automobilbau*. Düsseldorf, Germany: VDI Report No. 699, pp. 593–630.
60. Reik, W. (1990). Torsionsschwingungsisolation im Antriebsstrang—ein Wertungsversuch. In: *Torsionsschwingungen im Antriebsstrang*, Baden-Baden (pp. 119–138). Bühl, Germany: LuK.
61. Abel, A., Schreiber, U., Valsania, S., and Fornelli, A. (2005). Simulation-based design of gearboxes for high-performance sports cars. In: A. Laschet et al.: *Systemanalyse in der Kfz-Antriebstechnik III* (pp. 160–172). Renningen, Germany: Expert Verlag.
62. Schreiber, U. (2004). Modeling and simulation of the complete powertrain of the golf 4motion. *Internationale Zulieferbörse*. Wolfsburg, Germany: Volkswagen AG.
63. Abel, A., Schreiber, U., and Schindler, J. (2006). Engine and gearbox modeling and simulation for improving the shifting behavior of powertrains with manual or automated transmission. SAE Technical Paper Series, No. 2006-01-1641.

24 Brake System Dynamics

Carlo Maria Domenico Cantoni, Riccardo Cesarini, Giampiero R.M. Mastinu, Giorgio Previati, and Roberto Sicigliano

CONTENTS

24.1 Introduction 838
24.2 Brake Functions 838
24.3 Braking Processes 840
24.3.1 Stopping Distance: Service Braking 840
24.3.2 Vehicle at Standstill: Parking Brake 842
24.3.3 Straight Braking and Stability 844
24.3.4 Optimum Straight Braking and Brake Balance 846
24.3.4.1 Vehicles with More than Two Axles 849
24.3.4.2 Articulated Vehicles and Vehicle–Trailer Combinations 849
24.3.5 Vehicle Body Attitude When Braking 850
24.4 Hydraulic Brake System and Components 854
24.4.1 Brake System Configurations 858
24.4.2 Brake System Components 858
24.4.2.1 Pedal 859
24.4.2.2 Brake Booster 860
24.4.2.3 Tandem Master Cylinder 861
24.4.2.4 Pressure Limiters and Reducers 862
24.4.3 Brake Types 864
24.4.3.1 Drum Brakes 865
24.4.3.2 Disk Brakes 865
24.5 Retarders 866
24.6 Noise Vibration Harshness 867
24.6.1 NVH State-of-the-Art Reviews 868
24.6.2 Phenomena and Influencing Factors 869
24.6.3 Experimental and Computational Activities 871
24.6.3.1 Experimental Activities 871
24.6.3.2 Computational Activities 873
24.6.3.3 Both Experimental and Computational Activities 874
24.6.4 Models 874
24.6.4.1 Friction 874
24.6.4.2 Lumped-Parameter Models 878
24.6.4.3 Models Derived from a Modal Approach 890
24.6.4.4 FEM Models 892
24.6.5 Brake Vibration and/or Noise: Causes and Mechanisms 896
24.6.5.1 Judder 896
24.6.5.2 Groan, Moan, and Howl 898
24.6.5.3 Squeal Noise 898
24.7 Conclusion 900
References 902

24.1 INTRODUCTION

Brakes are the most important safety subsystem of each ground vehicle. As for other subsystems of the vehicle such as engine, suspensions, and tires, brakes are very complex engineering devices. Material science, applied mechanics and thermo-fluid dynamics, manufacturing technology, and other engineering disciplines are all necessary to conceive and manufacture an effective brake system.

Brakes were first invented by the Romans and were applied to carts or wagons (Figure 24.1). Dissipating into heat the kinetic energy was at that time (and still today) the more effective way to stop a vehicle.

In this chapter, an overview of the main brake functions and constructive applications is given. Additionally, brake comfort (noise–vibration–harshness [NVH]) is dealt with in the second part of the chapter.

24.2 BRAKE FUNCTIONS

In a motor vehicle, the brake system has three main functions [1]:

- Reduce speed and, eventually, stop the vehicle (service brake)
- Prevent unwanted acceleration, for instance when driving downhill (retarders)
- Keep the vehicle at standstill (parking brake)

For common vehicles powered by internal combustion engines, the kinetic energy is normally dissipated into heat; this is a very efficient way to produce braking forces [1]. For hybrid or electric vehicles, regenerative braking is obtained by exploiting electric motors [2]. In this chapter, only common dissipative brakes will be dealt with.

On horizontal roads, the energy dissipated during braking is given by the difference between the initial and final kinetic energy:

$$E_d = \frac{1}{2} m \left(v_i^2 - v_f^2 \right) + \frac{1}{2} J_{rot} \left(\omega_i^2 - \omega_f^2 \right) \tag{24.1}$$

where

m is the mass of the vehicle

v_i and v_f are, respectively, the initial and final velocity of the vehicle

J_{rot} is the global rotational inertia of tires, transmission, and engine (for the computation of J_{rot}, both the gear ratio of the gear box and the gear ratio of the final reduction have to be taken into account [3])

ω_i, ω_f are respectively, the initial and final rotational velocities of the wheels of the drive axle (for sake of simplicity, we deal with the case of one single driving axle)

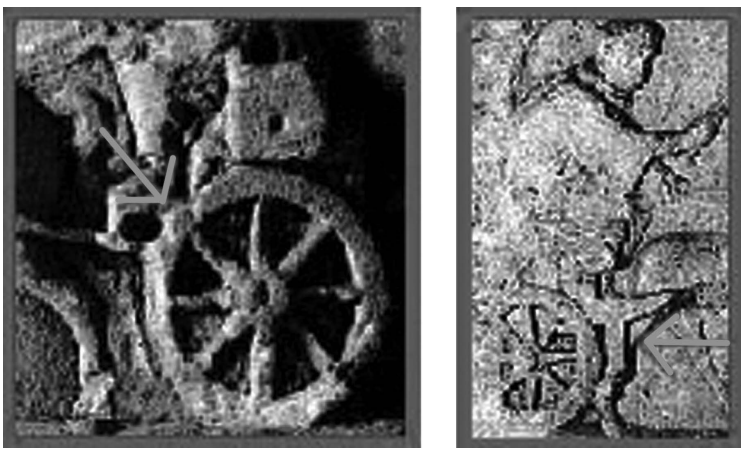


FIGURE 24.1 Ancient Roman carts or wagons with early brakes operated directly by the driver. (Adapted from Weller, J.A., Roman traction systems, www.humanist.de/rome/rts)

Often, the rotational energy is neglected, especially for non-driving axle(s). If other energy dissipation actions are neglected (aerodynamic drag forces, rolling resistance, engine resisting torque), all the kinetic energy must be dissipated by the brake system. In case of sloped roads, the potential energy of the gravitational field must be added to Equation 24.1:

$$E_d = \frac{1}{2} m \left(v_i^2 - v_f^2 \right) + \frac{1}{2} J_{rot} \left(\omega_i^2 - \omega_f^2 \right) - mg(h_f - h_i) \tag{24.2}$$

where

g is the gravitational acceleration

h_f and h_i are the initial and final height of the road

If $h_f > h_i$, the vehicle is braking uphill and part of its initial energy is turned into potential energy, thus reducing the energy dissipated by brakes. The more demanding situation is when $h_f < h_i$, the vehicle is braking downhill and the brake system has to dissipate both the kinetic and the potential energy. When driving downhill at a constant speed, the kinetic energy of the vehicle is constant; thus, the potential energy has to be completely dissipated by the brake system. In this case, retarders are employed [1,3].

The energy dissipated by the brake system is usually converted into heat that has to be evacuated from the brake system. The thermal resistance of the brake system is a crucial issue [1]. Moreover, different thermal behaviors are involved during speed reduction (service brake operation) or during driving downhill (retarder brake operation) [3]. In the first case, a great amount of thermal power is generated for a relatively small time. The thermal capacity of the brake accumulates the heat developed during braking and usually (for normal cars), a not very high-temperature increase occurs. In the second case, an almost constant thermal power is generated during the running of the vehicle at constant speed down a sloped path. The brake system is designed to reach a thermal equilibrium up to the admissible temperature limit of the components, such temperatures are generally much higher than those reached during service brake operation [1,3].

An estimation of the thermal power that the brakes have to dissipate when braking on a horizontal road can be given as follows. Considering an acceleration–braking cycle, the mean thermal power in the cycle can be computed as

$$W_d = \frac{1}{2} \frac{mv^2 + J_{rot}\omega^2}{t_a + t_b} \tag{24.3}$$

where t_a and t_b are the times needed to accelerate the vehicle from standstill to v and to brake from v to standstill, respectively. The mean power is null either when the velocity is null or when the velocity is equal to the maximum speed of the vehicle (in this case, $t_a \to \infty$ as the maximum speed is ideally an asymptotic value). A value of v can be computed for which the mean power reaches a maximum. This value depends on the mass of the vehicle and on its performance (performances of both engine and brakes).

The thermal power when driving downhill can be computed as

$$W_d = mgv \sin(\alpha) \tag{24.4}$$

α being the slope angle of the road. For a given slope angle, the thermal power linearly increases with the velocity. A limit velocity is imposed, at each slope angle, by the thermal dissipation performance of the brake system.

Usually, for light high-performance vehicles, the most critical condition is given by the service brake situation. For heavy commercial vehicles, the downhill situation is the most critical. For these vehicles, the friction brakes are devoted to the service brake and a dedicated subsystem realizing the retarder function is needed. Figure 24.2 shows the thermal power dissipated for level and sloped roads [4,5].

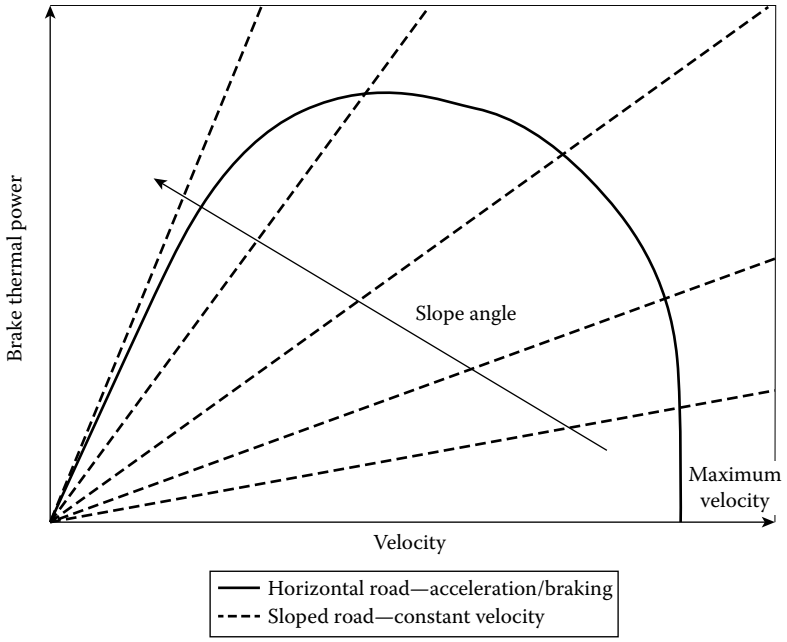


FIGURE 24.2 Brake thermal power dissipated as function of velocity for horizontal road (continuum line) and sloped road (dashed line).

24.3 BRAKING PROCESSES

24.3.1 Stopping Distance: Service Braking

The brake system has the task, especially in emergencies, of stopping the vehicle in the shortest possible distance. The stopping distance, however, does not depend on the braking system alone, but the following factors have to be considered.

Figure 24.3 shows the braking process over time [1,6,7]. At time zero, the driver sees an obstacle. After the reaction time t_r, the driver starts to apply the pedal force in order to brake the vehicle. The reaction time depends on many factors, among which driver attention and age play a very important role. The estimation of the reaction time is a very complex operation; a commonly accepted average value is around 1 s. In the figure, the pedal force is simplified as a linear function of time. The actual shape of the pedal force depends on the ability of the driver. More-skilled drivers are able to apply a steady rising pedal force, while unskilled drivers present strong oscillation in the pedal force.

After the reaction time, some more time has to be considered before vehicle deceleration begins due to the time needed by the braking system to start applying the braking force, that is, clearance taking up. This time is called brake application time (t_a) and is in the order of 0.1–0.2 s for hydraulic braking systems and up to 1 s for pneumatic braking systems [6,7].

In this first part of the maneuver (area 1 in the figure), with a total duration of $t_r + t_a$, the deceleration can be considered null; only a small deceleration could be present due to aerodynamic and motor drag but can be neglected [6]. The speed of the vehicle is constant and equal to the speed before the driver has seen the obstacle. The distance traveled in this phase can be of the order (if not greater) of the distance needed for the actual braking of the vehicle.

In phase 2 of Figure 24.3, the vehicle deceleration grows according to a law depending on the application of the pedal force and on the dynamics of the braking system. The maximum deceleration is reached, either because of tire locking or because the pedal force is kept constant. In this phase, vehicle speed begins to reduce; the deceleration can be considered a linear function of

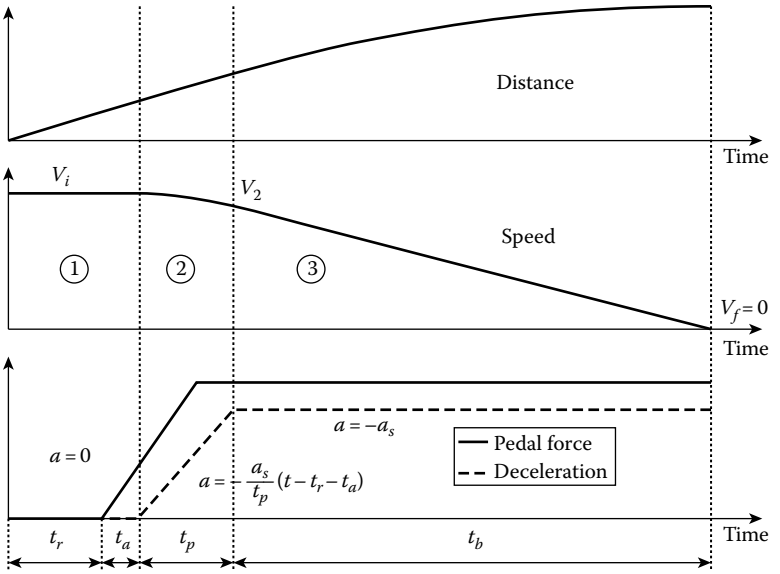


FIGURE 24.3 The braking process over time.

the time up to the stabilized value of deceleration a_s. At a_s (phase 3), the speed decreases linearly with time till the vehicle stops.

The total stopping distance is given by the sum of the distance traveled in the three different phases:

Phase 1: constant velocity motion, the traveled distance s_1 can be computed as

$$s_1 = V_i \cdot (t_r + t_a) \tag{24.5}$$

where V_i is the initial velocity of the vehicle. The velocity at the end of this phase is still V_i.

Phase 2: motion with linear rising deceleration. Considering a linear rising deceleration between zero and the stabilized value a_s in the time t_p, the traveled distance and the velocity variation in this phase can be estimated as

$$\begin{cases} a = -\dfrac{a_s}{t_p} \cdot t \\ V_2 = V_i - \displaystyle\int_0^{t_p} \frac{a_s}{t_p} t dt = V_i - \frac{1}{2} a_s t_p \\ s_2 = V_i \cdot t_p - \displaystyle\int_0^{t_p} \frac{1}{2} \frac{a_s}{t_p} t^2 dt = V_i \cdot t_p - \frac{1}{6} a_s t_p^2 \end{cases} \tag{24.6}$$

Phase 3: motion with constant deceleration, the traveled distance and the velocity are

$$\begin{cases} a = -a_s \\ V_f = V_2 - a_s t_b \\ s_3 = V_2 \cdot t_s - \dfrac{1}{2} a_s t_b^2 \end{cases} \tag{24.7}$$

where t_b is the time to stop the vehicle and can be computed by imposing $V_f = 0$. The resulting traveled distance is

$$s_3 = \frac{1}{2}\frac{V_2^2}{a_s} \tag{24.8}$$

when $t_b = V_2/a_s$.

By replacing the second of Equation 24.7 into Equation 24.8 and summing all of the three contribution, the total stopping distance reads

$$s = s_1 + s_2 + s_3 = V_i\left(t_r + t_a + \frac{t_p}{2}\right) + \frac{1}{2}\frac{V_i^2}{a_s} - \frac{1}{24}a_s t_p^2 \tag{24.9}$$

The last term of Equation 24.9 is usually small with respect to the other terms. Referring to a braking maneuver from 100 km/h, with $t_r = 1$ s, $t_a = 0.2$ s, $t_p = 0.3$ s, and $a_s = 8$ m/s^2, the total braking distance s is 85.3 m and the last term of the equation is less than 0.1 m.

A simpler expression for the estimation of the stopping distance can be obtained by considering only a phase with no deceleration and a phase with constant deceleration. To consider the time needed for the brake system to rise the braking force, a rising time to reach the 75% of the braking torque can be added to the reaction time. In this case, the stopping distance can be estimated as

$$s = \frac{1}{2}\cdot\frac{V_i^2}{a_s} + V_i \cdot t_{r_tot} \tag{24.10}$$

where

V_i is the initial velocity before braking
a_s is the constant deceleration during braking
t_{r_tot} is the total reaction time accounting for driver reaction time, brake activation time, and rising time

In case tire–road friction is fully exploited, the deceleration can be considered as $a_s = \mu g$, μ being the tire–road friction coefficient and g the gravity. Equation 24.10 can be rewritten as a function of the tire–road friction coefficient as

$$s = \frac{1}{2}\cdot\frac{V^2}{\mu g} + V \cdot t_{r_tot} \tag{24.11}$$

In Figure 24.4a, the respective total braking distances for decelerations on a high-friction road ($\mu = 0.8$) or on a low-friction road ($\mu = 0.4$) are reported as a function of the initial speed (ranging from 20 to 200 km/h) along with the distance traveled during the total reaction time (considered as 1.5 s). In Figure 24.4b, the percentage of the total braking distance traveled during the total reaction time is depicted for the two cases. In case of relatively small velocities and high road friction, the most part of the braking distance is traveled during the total reaction time, when the driver is not able to make any real action to prevent accidents (for more details on braking distances, see [1,6,8]).

24.3.2 Vehicle at Standstill: Parking Brake

In this section, a vehicle at standstill on a sloped road is considered. The brake system has to apply the braking force necessary to keep still the vehicle. Referring to Figure 24.5, the equilibrium of the vehicle can be written as

$$\begin{cases} mg\cos(\alpha) - F_{zF} - F_{zR} = 0 \\ mg\sin(\alpha) - F_{xF} - F_{xR} = 0 \end{cases} \tag{24.12}$$

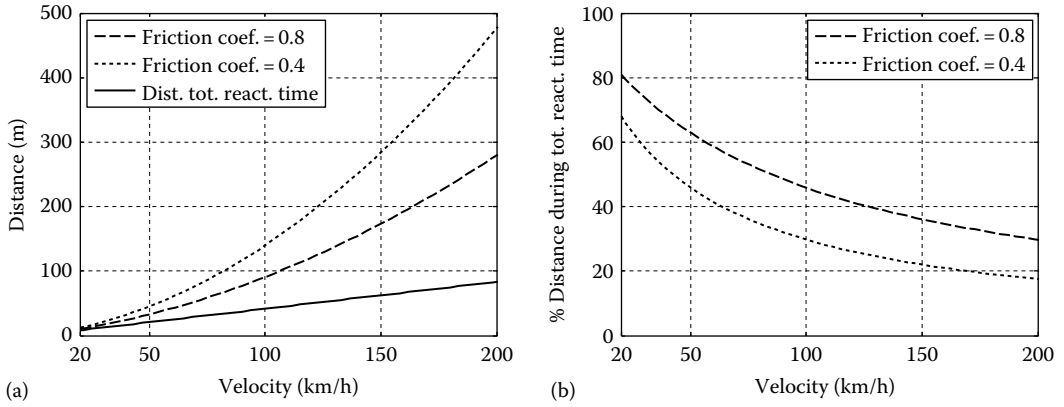


FIGURE 24.4 (a) Total braking distance for a high-friction and a low-friction road and (b) Percentage of the total distance traveled during the total reaction time in case of high-friction and of low-friction road.

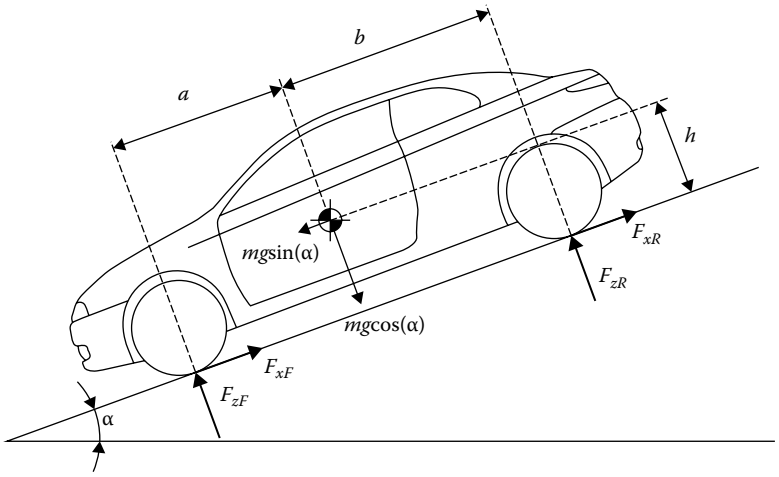


FIGURE 24.5 Braked vehicle standing on a sloped road.

If the tires have all the same friction coefficient μ, the limit condition to prevent sliding of the car along the slope is

$$mg\sin(\alpha) \le F_{xF} + F_{xR} = \mu(F_{zF} + F_{zR}) = \mu(mg\cos(\alpha)) \tag{24.13}$$

and the limit slope angle results

$$\tan(\alpha_{lim}) = \mu \tag{24.14}$$

For slope angles $\alpha > \alpha_{lim}$, the vehicle cannot be hold at standstill but it will slide downhill. In most cases, the parking brake acts on the rear axle only. In this situation, $F_{xF} = 0$ and the limit slope angle can be computed as

$$\begin{cases} mg\cos(\alpha) - F_{zF} - F_{zR} = 0 \\ mg\sin(\alpha) - F_{xR} = 0 \end{cases} \tag{24.15}$$

$$mg\sin(\alpha) \le F_{xR} = \mu F_{zR} = \mu mg\frac{a\cos(\alpha) - h\sin(\alpha)}{a+b} \tag{24.16}$$

And the resulting limit slope angle

$$\tan(\alpha_{lim}) = \mu \frac{a}{a + b + \mu h} \tag{24.17}$$

(For more details on such topic, see [9].)

24.3.3 Straight Braking and Stability

In order to get the best possible performance from a brake system, that is, achieving the shorter stopping distance, all of the available road–tire friction of each tire should be exploited. This means that if all of the tires have the same friction coefficient, the braking torque should be distributed to the tires according to the actual vertical load acting on each tire. Considering a symmetric vehicle, the load on tires on the same axle is the same and the braking torque should be distributed according to the load transfer from the rear axle to the front axle during braking. If the brake torque rises during braking, one of the following conditions occurs:

- The rear tires lock first.
- The front tires lock first.

These two situations have very different effects on vehicle stability. A quasi-static simplified analysis of the two scenarios can be carried out as follows [6,10]. Figure 24.6 depicts the two situations of front (left) and rear (right) locked tires. If the rear tires are locked first, the rear tire can react with a force opposite to the sliding velocity vector. If a disturbing lateral force is applied to the center of gravity of the vehicle, a slip angle appears between the longitudinal axis of the vehicle and the direction in which the center of gravity is moving. A reacting force can arise only at the front axle, thus generating a yaw moment that increases the slip angle of the vehicle, that is, the longitudinal axis of the vehicle rotates as to increment its angle with respect to the velocity direction. The system of forces is thus diverging and tends to increase the slip angle, which implies a very difficult handling of the vehicle. Only very skilled, and professional, drivers could control a vehicle in such a situation. A normal driver will lose control of the vehicle and very dangerous accidents can happen.

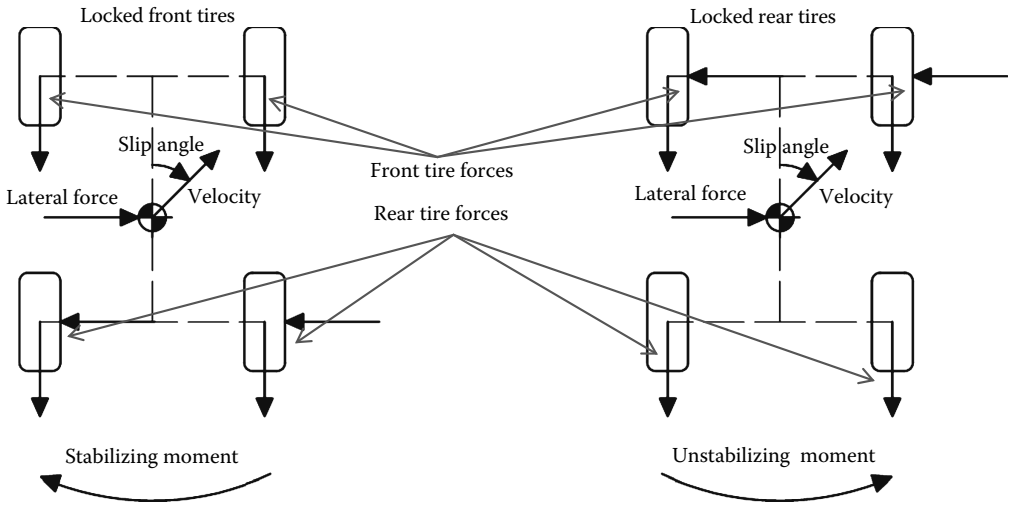


FIGURE 24.6 Vehicle behavior with rear and front wheels locked.

If the front tires lock first, the reaction force appears at the rear axle and the consequent yaw moment aligns the vehicle to the velocity vector of the center of gravity, thus reducing the slip angle of the vehicle. The resulting system of forces is converging and the handling of the vehicle is good.

A deeper understanding of the dynamics of the vehicle when the tire locking is reached can be achieved by considering the so-called *handling diagrams*. The theory leading to the definition and computation of the handling diagrams has been described in Sections 11.4 and Section 27.3, where the most important considerations on vehicle stability have been explained into details. A very important result of such discussions is that oversteering vehicles become unstable for velocities over a certain critical velocity. The instability can occur even on a straight road. The characterization of a vehicle as understeering or oversteering is given according to the understeering gradient and can be defined as (see Equation 11.47 or Equation 26.88)

$$K_{US} = \frac{m(C_R b - C_F a)}{(a+b)C_R C_F} \tag{24.18}$$

where

m is the mass of the vehicle

C_F and C_R are front and rear *cornering* stiffness, respectively

a and b are the dimensions of the vehicle as shown in Figure 24.5

A vehicle is said to be oversteering when $K_{US} > 0$ that implies $C_R b < C_F a$.

When a tire is locked, the lateral force is rather small, but not vanishing; it can be computed as $F_y = \mu F_z sin(\alpha)$ [11]. Figure 24.7 shows the handling diagrams for a vehicle either with all tires rolling or with locked front tires or with locked rear tires. Also, the characteristics of the unlocked axles are much lower than their pure lateral characteristics due to the presence of the braking force in longitudinal direction [11]. Let us consider rolling tires (Figure 24.7a), every vehicle is designed to be understeering. A vehicle with locked front tires (Figure 24.7b) is understeering, while a vehicle with locked rear tires (Figure 24.7c) is oversteering (see Chapter11). In fact, the cornering stiffness (C_α) for small values of α for locked tires is $C_\alpha = \mu F_z$ ($\mu = 0.6/1.2$ for dry pavements) instead $C_\alpha/F_z = 10/25$ for rolling tires ($C_\alpha/F_z = 4/7$ for truck tires) [3]. Thus, if the rear tires are locked, K_{US} (Equation 24.18) becomes negative. In [11], the interested reader can find the full discussion of this topic along with the computation of the critical velocity for a vehicle with locked rear tires.

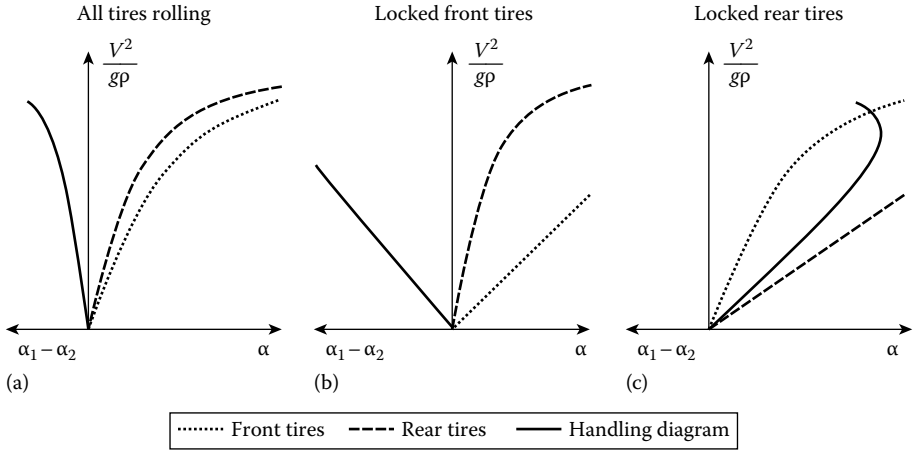


FIGURE 24.7 Handling diagrams (for the computation of the diagrams, see Section 11.4). Locked front tires produce high understeering; locked rear tires produce high oversteering and possibly instability even at straight driving. a: all tires rolling. b: locked front tires. c: locked rear tires.

The brake system has to be designed in order to lock the front tires first. In this case, if the locking of the front tires is reached, the vehicle will slide straight without responding to steering inputs by the driver. If a collision is unavoidable, a frontal crash typically will result in fewer injuries than a side crash due to better occupant protection.

24.3.4 Optimum Straight Braking and Brake Balance

In Chapter 11, the fundamental relationships between front and rear braking forces to have the optimal braking balance have been derived and discussed. Referring to Figure 24.8, these relationships can be written as

$$\frac{F_{xF}}{mg} = \frac{\mu}{L} \cdot \left(b + \frac{a_x}{g} h \right) \tag{24.19}$$

$$\frac{F_{xR}}{mg} = \frac{\mu}{L} \cdot \left(a - \frac{a_x}{g} h \right) \tag{24.20}$$

where μ is the friction coefficient. The loads on the front and rear axles depend on the actual value of acceleration and on the position of the center of gravity

$$\frac{F_{xR}^2}{m^2 g^2} + 2\frac{F_{xR}F_{xF}}{m^2 g^2} + \frac{b}{h}\frac{F_{xR}}{mg} - \frac{a}{h}\frac{F_{xF}}{mg} + \frac{F_{xF}^2}{m^2 g^2} = 0 \tag{24.21}$$

Equation 24.21 is a parabola and represents the ideal ratio between front and rear longitudinal forces to have optimum braking if the friction is equal at front and rear tires. In Figure 24.9, this relationship is depicted for a laden and an unladen vehicle [1–6,12,13]. The points on the parabola represent the front and rear forces to have the simultaneous locking of all of the tires. At the points located under the parabola, the front tires lock first and at the points located over the parabola, the rear tires lock first (see Figure 24.9).

The design corridor for the brake distribution can be found as the area between the most critical condition to avoid rear tire locking (usually unladen vehicle configuration) and the maximum acceptable braking distance (for Europe, this distance is defined by the Union Directive 71/320 [14]). For a typical hydraulic brake system employed on cars, the ratio between front and rear braking forces is fixed and it is depicted as a straight line in Figure 24.9 (fixed brake force distribution [BFD]). It appears that a fixed BFD is not effective in obtaining high braking performances. In fact, at low longitudinal acceleration, such a distribution gives much smaller forces than the ideal distribution.

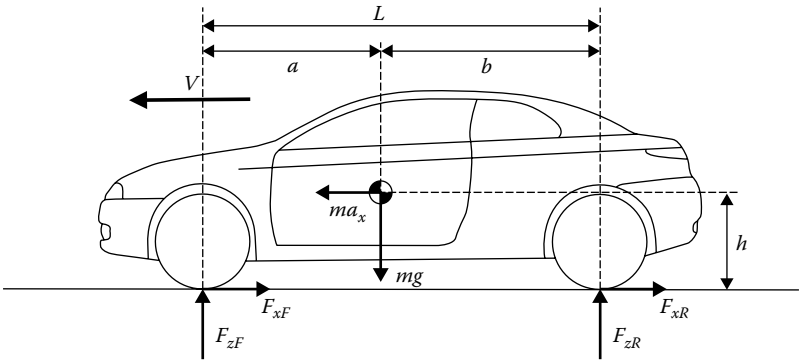


FIGURE 24.8 Vehicle braking on a horizontal road.

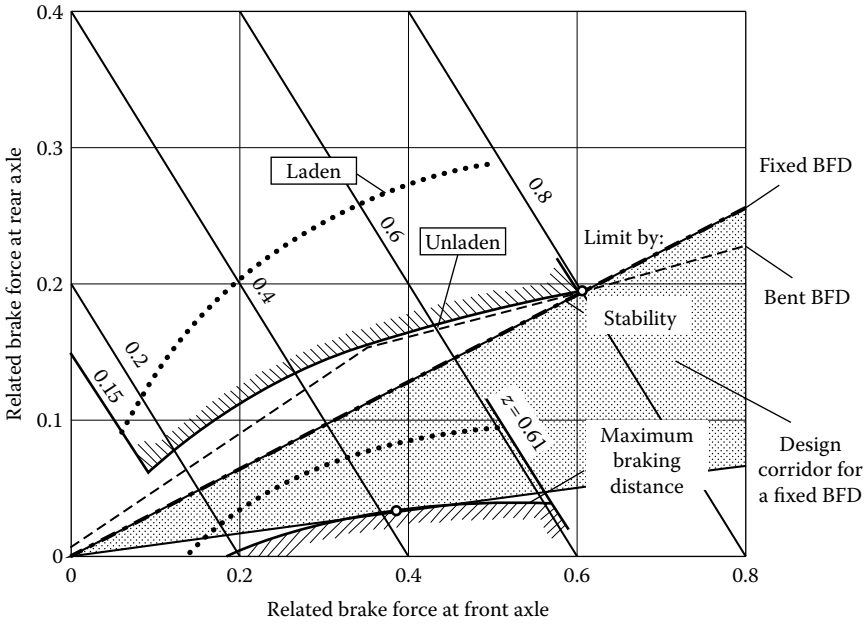


FIGURE 24.9 Fixed, bent, and load-sensitive brake force distribution (BFD). (Adapted from Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, 2008, ISBN 978-0-7680-1787-8.)

At high longitudinal acceleration, the vehicle tends to lock the rear tires first. Better results are obtained with a bent BFD in which a limit valve is employed for limiting the rear brake pressure at high longitudinal acceleration. In this way, a more convenient balance can be used at low longitudinal acceleration and the locking of the rear tires can be avoided at high longitudinal accelerations. In case of a laden vehicle, the brake system performances can be improved by varying the limit valve operation point as to allow for higher rear braking forces (load-sensitive limit valve, see Section 24.4.2.4).

In a hydraulic brake system without proportioning valve, the pressure is equal at the front and rear axles; thus, a fixed ratio exists between the front and rear brake forces and depends on the different brakes installed on the two axles. The brake distribution (φ), or brake balance, of a vehicle can be defined as the ratio between the rear braking force and the total braking force

$$\phi = \frac{F_{xR}}{F_{xF} + F_{xR}} \tag{24.22}$$

F_{xF} and F_{xR} being the braking forces at the front and rear axle, respectively. The best compromise brake distribution is obtained by proper design of the front and rear brakes. As shown in Figure 24.9, a fixed brake balance is not an optimal choice for the brake system. To improve the brake performance at low accelerations, a pressure-limiting or pressure-reducing valve is employed at the rear axle. At a given pressure, the valve intervenes either limiting or reducing the rear pressure, changing the slope of the balance curve.

It should be noted that the transmission drag and rotary inertia have practical interest when designing the BFD of the vehicle. The effect of the transmission drag is to add a braking torque generated to the driving axles. The effect of the rotary inertia is to use part of the brake torque to reduce the rotational speed of the rotating parts. Both driving and not driving wheels are affected by this phenomenon, but it is more relevant for driving axle where the rotational velocity of wheels and transmission has to be reduced.

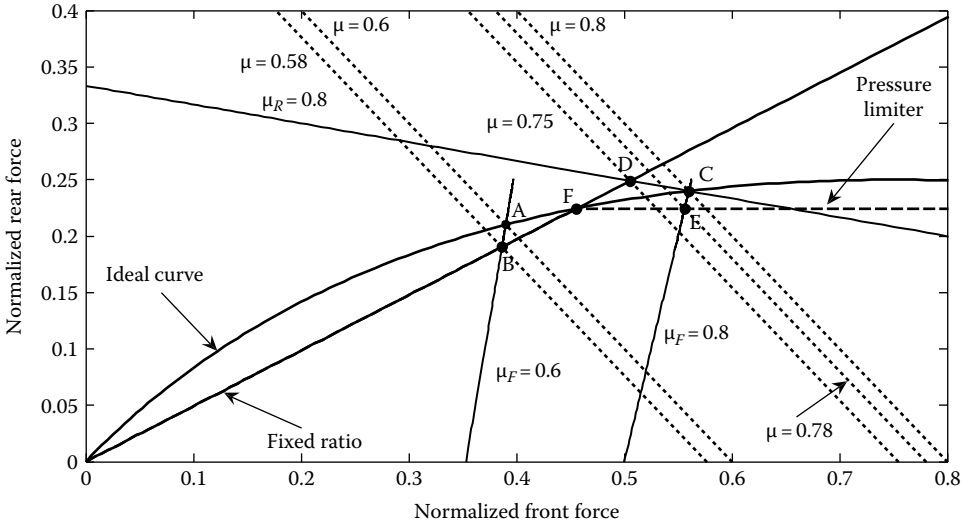


FIGURE 24.10 Optimum and real brake balance.

Aerodynamic effects on the brake balance should be considered when aerodynamic foils are employed on the vehicle (see Chapter 15). For standard vehicles, the effect of aerodynamic force is negligible in most practical situations.

Figure 24.10 shows the situation of a braking system with a fixed brake balance. If a friction coefficient of 0.6 is considered at all tires, point A represents the ideal brake distribution to obtain a deceleration of 0.6 g. If a fixed brake balance is considered as depicted in the figure, for this level of friction coefficient, the front tires lock first. The brake forces, front and rear, increase during pedal force application along the line representing the fixed brake balance up to point B. At this point, the fixed balance line crosses the line representing a constant front friction of 0.6 (for the definition of the lines at fixed front or rear friction, see Figure 11.30) and the front tires approach locking condition with a vehicle deceleration equal to 0.58 g. If more pedal force is applied, point A is reached with the front tires locked and the increase in the front force is due to load transfer.

If a friction coefficient of 0.8 is considered, for the depicted vehicle, the rear tires lock first. Point C represents the ideal brake balance to obtain a deceleration of 0.8 g. The actual brake forces follow the line representing the fixed brake balance up to point D where there is the intersection with the line representing a fixed rear friction coefficient of 0.8. At this point, the vehicle deceleration is equal to 0.75 g and the rear tires approach locking condition. The vehicle is in an unstable braking situation (continuing depressing the brake pedal from point D, point C would be reached; at C, both of the front and rear tires are locked). To avoid the locking of rear tires, a pressure-limiting or reducing valve can be employed. If a pressure-limiting valve is employed with a cut pressure equal to the rear pressure at point F, the brake forces have to increase according to the dashed line of figure from point F to point E where there is the intersection with the line of constant front friction of 0.8. At this point, the front tires approach locking condition at a deceleration of 0.78 g with a stable braking configuration.

The brake efficiency (E) can be defined as the ratio between the actual longitudinal deceleration a_x and $\mu \cdot g$

$$E = \frac{a_x}{\mu g} \tag{24.23}$$

Similarly, E can be defined for the front and rear axle as the ratio between the longitudinal deceleration and the friction coefficient at that axle times g. The friction coefficients at the front (μ_F) and rear (μ_R) axle can be computed as the ratio between the actual longitudinal force and vertical force

at that axle. For a fixed brake balance, with the notation of Figure 24.8, the front and rear friction coefficients can be computed as [6]

$$\mu_F = \frac{F_{xF}}{F_{zF}} = \frac{(1-\phi)(a_x/g)}{(b/L)+(a_x/g)(h/L)} \tag{24.24}$$

$$\mu_R = \frac{F_{xR}}{F_{zR}} = \frac{\phi(a_x/g)}{(a/L)-(a_x/g)(h/L)} \tag{24.25}$$

$$\text{beings with } F_{xF} = (1-\phi)\cdot a_x m, \quad F_{zF} = mg\left(\frac{b}{L}+\frac{a_x}{g}\frac{h}{L}\right), \quad F_{xR} = \phi\cdot a_x m,$$

$$\text{and } F_{zF} = mg\left(\frac{a}{L}-\frac{a_x}{g}\frac{h}{L}\right)$$

The front and rear brake efficiencies read

$$E_F = \frac{a_x}{\mu_F g} = \frac{(b/L)}{1-\phi-(h/L)\mu_F} \tag{24.26}$$

$$E_R = \frac{a_x}{\mu_R g} = \frac{(a/L)}{\phi+(h/L)\mu_R} \tag{24.27}$$

Only values smaller than 1 are physically meaningful.

24.3.4.1 Vehicles with More than Two Axles

The analysis of brake dynamics of vehicle with more than two axles is important to define the correct brake balance between the axles of the vehicle. Typical vehicles with more than two axles are three-axle trucks or buses with a single front axle and a tandem rear axle. For this type of vehicles, the dynamic equilibrium of the tandem rear axle has to be considered in order to compute the load transfer between the three axles. The equation involved is generally complex and computer programs can be employed for the solution. In some cases, such as walking beam or leaf spring tandem axle suspensions, the equation can be solved in a closed form [6].

In general, the load transfer increases the load on the front axle and reduces the load on the rearmost axle. The load on the central axle has smaller variations. The brake system has to be realized in order to optimize the brake torque application as a function of the vertical load variation.

24.3.4.2 Articulated Vehicles and Vehicle–Trailer Combinations

In case of articulated vehicles or vehicle–trailer combinations (Figure 24.11), the brake balance is crucial for stability [3,6]. Figure 24.11a refers to a tractor and trailer combination. The trailer has two axles and can support its own weight. Referring to the symbols in the figure, just formally, the respective decelerations of the tractor (a) and of the trailer (a_t) can be denoted as

$$a = \frac{F_x}{mg} \quad a_t = \frac{F_{xt}}{m_t g} \tag{24.28}$$

The deceleration (a_s) of the tractor and trailer combination can be expressed as

$$a_s = a\frac{mg}{mg+m_t g} + a_t\frac{m_t g}{mg+m_t g} \tag{24.29}$$

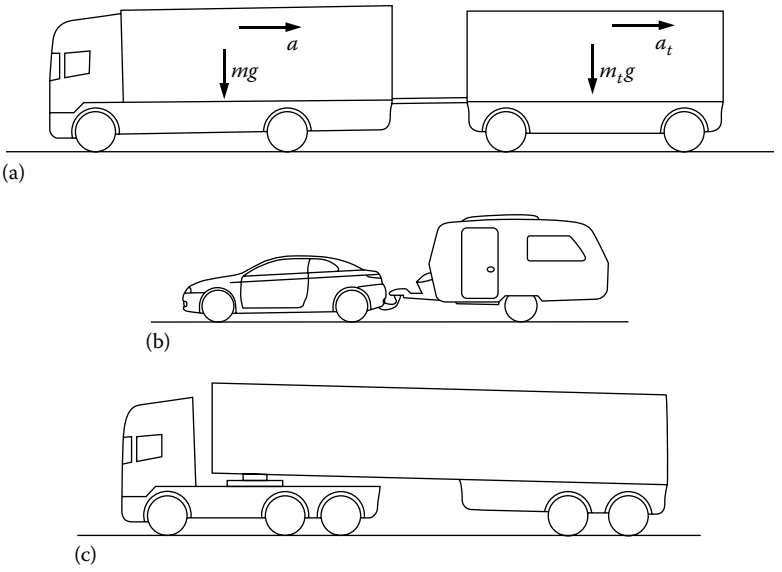


FIGURE 24.11 Articulated vehicles: (a) tractor and trailer, (b) car and caravan, and (c) tractor and semitrailer.

Two dangerous situations can arise. If the deceleration of the tractor is higher than the deceleration of the trailer ($a > a_t$), the trailer exerts a horizontal force on the tractor in the direction of the motion. The link between the tractor and the trailer is compressed and an instability may occur. In case the deceleration of the trailer is higher than the deceleration of the tractor ($a_t > a$), there is a risk of locking of the tires of the trailer. In this case, the trailer is no longer able to follow the tractor during curve negotiation, that is, the trailer can slide outside the curve [3]. To avoid these two opposite situations, the tractor and trailer must have the same deceleration, that is, forces F_x and F_{xt} must be tuned in order to have $a = a_t$. The horizontal force between tractor and trailer has to vanish during braking.

Figure 24.11b shows an articulated vehicle realized by a car and a caravan. In this case, the caravan is not able to support its own weight and a vertical force is acting at the hitch. Also, a certain amount of horizontal force cannot be avoided. The forces acting between car and caravan change the vertical loads acting at the front and rear axles of the car. In particular, the vertical load at the rear axle is increased, while the vertical load at the front axle is decreased. The brake balance of the car has been set in order to have the best performances when the car is riding alone. In this case, the balance should be shifted toward the rear axle. If this cannot be done, bad performances are to be expected. If an electronic system is present, the brake balance can be adjusted for this situation. Moreover, if the caravan does not have a braking system, the system can become unstable in a similar way of the previous case when the trailer has less deceleration of the tractor. If the caravan can brake, given the differences in the adherence of the tires of the caravan and of the car, there is the risk of locking the tires of the caravan [3].

A tractor and semitrailer combination is depicted in Figure 24.11c. This situation is very similar to the car–caravan situation. However, the risks of instability are limited by the positioning of the connecting hitch between the axles of the tractor, by designing the braking systems of the tractor and of the trailer as to work together, and by the presence of a mandatory electronic system for the regulation of the braking force balance between the tractor and trailer. (More information on this topic can be found in [15,16].)

24.3.5 Vehicle Body Attitude When Braking

In conventional vehicles, with front and rear suspension systems, the body pitch attitude during braking or acceleration varies due to the effects of the longitudinal acceleration. In general, important pitch of the body occurs, which is strongly related to the geometry of the suspension system. This

section is devoted to the analysis of the forces transmitted by front and rear suspensions to the body of the vehicle in order to derive the geometrical features of the suspension system required to avoid the pitching of the body. The feature of the suspension system that prevents the pitching of the car body when braking is called anti-dive characteristic. A suspension system that totally avoids brake pitch rotation has a 100% anti-dive.

For a car with two axles and brakes located on the wheels, following the comprehensive work reported in [3], the braking dive phenomenon can be described as follows.

In Figure 24.12, a suitable model for the definition of the anti-dive characteristics of a suspension system is depicted. The model has 3 degrees of freedom (DOFs), namely, longitudinal displacement x, vertical displacement z, and pitch rotation φ. The front and rear elastokinematics are described by equivalent stiffness and damping and additionally by equivalent longitudinal arms with given inclinations. By such a model, every kind of suspension systems can be taken into account. The brakes are located at the wheels hub carriers, that is, the reaction torques of the brakes act directly at the hub carriers. Referring to the symbols reported in the figure, the dynamic equilibrium while braking reads

$$ma_x = -F_{xcf} - F_{xcr} \tag{24.30}$$

$$m\ddot{z} = F_{zcf} + F_{zcr} + F_f + F_r \tag{24.31}$$

$$J\ddot{\varphi} = -F_{zcf}(a - L_f) + F_{zcr}(b - L_r) - F_f a + F_r b - F_{xcf}(h - h_f) - F_{xcr}(h - h_r) \tag{24.32}$$

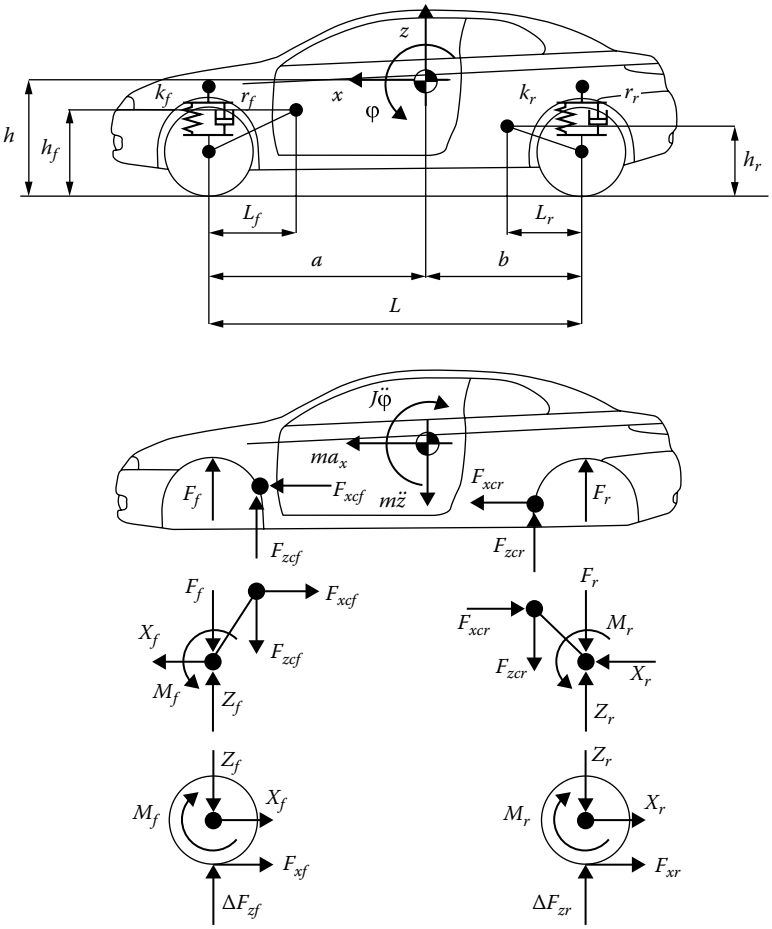


FIGURE 24.12 Vehicle model for anti-dive characteristic computation.

m being the mass of the body of the vehicle and J its moment of inertia with respect to the pitch axle. For the (massless) suspensions arms,

$$F_{xcf} = X_f \qquad F_{xcr} = X_r \tag{24.33}$$

$$F_{zcf} = Z_f - F_f \qquad F_{zcr} = Z_r - F_r \tag{24.34}$$

$$F_{zcf} L_f - F_{xcf}(h_f - r) - M_f = 0 \qquad F_{zcr} L_r + F_{xcr}(h_r - r) - M_r = 0 \tag{24.35}$$

r being the tire radius. For the (massless) tires,

$$F_{xf} = -X_f \qquad F_{xr} = -X_r \tag{24.36}$$

$$\Delta F_{zf} = Z_f \qquad \Delta F_{zr} = Z_r \tag{24.37}$$

$$M_f = F_{xf} r \qquad M_r = F_{xr} r \tag{24.38}$$

The forces due suspension stiffness and damping can be computed as

$$F_f = -k_f(z - L_f\varphi) - r_f(\dot{z} - L_f\dot{\varphi}) \tag{24.39}$$

$$F_r = -k_r(z + L_f\varphi) - r_r(\dot{z} + L_r\dot{\varphi}) \tag{24.40}$$

From Equations 24.30, 24.33, and 24.36,

$$ma_x = -F_{xcf} - F_{xcr} = X_f + X_r = F_{xf} + F_{xr} \tag{24.41}$$

By replacing Equation 24.41 into Equations 24.31 and 24.32, remembering Equations 24.34, 24.35, and 24.37 through 24.40, and rearranging, the vertical and pitch dynamic equilibrium of vehicle body can be rewritten as

$$m\ddot{z} + (r_f + r_r)\dot{z} - (r_f L_f - r_r L_r)\dot{\varphi} + (k_f + k_r)z - (k_f L_f - k_r L_r)\varphi = ma_x\left(\frac{F_{xf}}{F_{xf} + F_{xr}}\frac{h_f}{L_f} - \frac{F_{xr}}{F_{xf} + F_{xr}}\frac{h_r}{L_r}\right) \tag{24.42}$$

$$J\ddot{\varphi} - (r_f L_f - r_r L_r)\dot{z} + \left(r_f L_f^2 + r_r L_r^2\right)\dot{\varphi} - (k_f L_f - k_r L_r)z + \left(k_f L_f^2 + k_r L_r^2\right)\varphi = ma_x\left(h - \frac{F_{xf}}{F_{xf} + F_{xr}}\frac{h_f}{L_f}a - \frac{F_{xr}}{F_{xf} + F_{xr}}\frac{h_r}{L_r}b\right) \tag{24.43}$$

If the respective right expressions of Equations 24.42 and 24.43 were to vanish, both the heave (z) and the pitch (φ) would vanish. This occurs if

$$\frac{h_f}{L_f} = \frac{F_{xf} + F_{xr}}{F_{xf}}\frac{h}{L} \tag{24.44}$$

$$\frac{h_r}{L_r} = \frac{F_{xf} + F_{xr}}{F_{xr}}\frac{h}{L} \tag{24.45}$$

Please notice that if Equations 24.44 and 24.45 hold, no heave (z) and pitch (φ) occur both during steady-state braking and during transients, no matter which brake history is applied [3]!

If the steady-state equilibrium is considered, Equations 24.42 and 24.43 read

$$(k_f + k_r)z - (k_f L_f - k_r L_r)\varphi = ma_x\left(\frac{F_{xf}}{F_{xf}+F_{xr}}\frac{h_f}{L_f} - \frac{F_{xr}}{F_{xf}+F_{xr}}\frac{h_r}{L_r}\right) \tag{24.46}$$

$$-(k_f L_f - k_r L_r)z + \left(k_f L_f^2 + k_r L_r^2\right)\varphi = ma_x\left(h - \frac{F_{xf}}{F_{xf}+F_{xr}}\frac{h_f}{L_f}a - \frac{F_{xr}}{F_{xf}+F_{xr}}\frac{h_r}{L_r}b\right) \tag{24.47}$$

Solving the system of Equations 24.46 and 24.47, the steady-state pitch angle φ can be computed as

$$\varphi = \frac{ma_x}{L}\left(\frac{h}{L}\left(\frac{1}{k_f}+\frac{1}{k_r}\right) - \frac{F_{xf}}{F_{xf}+F_{xr}}\frac{h_f}{L_f}\frac{1}{k_f} - \frac{F_{xr}}{F_{xf}+F_{xr}}\frac{h_r}{L_r}\frac{1}{k_r}\right) \tag{24.48}$$

Figure 24.13 gives a graphical representation of the previous conditions for 100% anti-dive. The pitch rotation center, which can be obtained by the intersection of the lines passing through the tire contact points and the hinges between the suspensions arms and the car body, has the same height of the center of gravity. In this way, the inertia forces due to the deceleration pass through the pitch rotation center and no pitch moment is applied to the car body.

In case the brakes are located on the chassis of the car, the previous formulation has to be modified. In this case, the brake reaction torques act on the car body. The anti-dive conditions of Equations 24.47 and 24.48 can be recomputed and read

$$\frac{h_f}{L_f} = \frac{F_{xf}+F_{xr}}{F_{xf}}\frac{h}{L} + \frac{r}{L_f} \tag{24.49}$$

$$\frac{h_r}{L_r} = \frac{F_{xf}+F_{xr}}{F_{xr}}\frac{h}{L} + \frac{r}{L_r} \tag{24.50}$$

The graphical representation of the 100% anti-dive conditions for the case of brakes located on the car body is depicted in Figure 24.14. In this case, the pitch rotation center can be found by intersecting the lines passing through the center of the tires and the suspension arm hinges with the chassis and it is above the center of gravity. The consequent moment due to the inertial force during braking is equal to the reaction moment that the brakes apply to the car body.

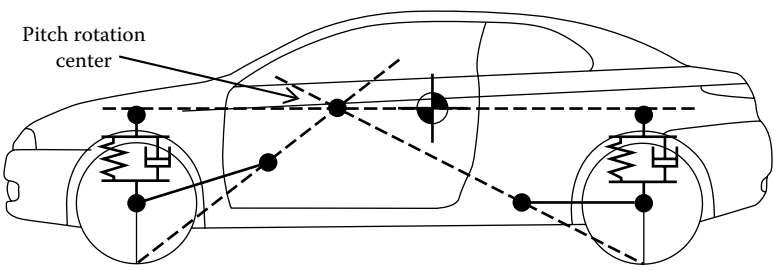


FIGURE 24.13 Front and rear suspension arms inclinations to have a 100% anti-dive configuration (brakes reacting on suspension arms).

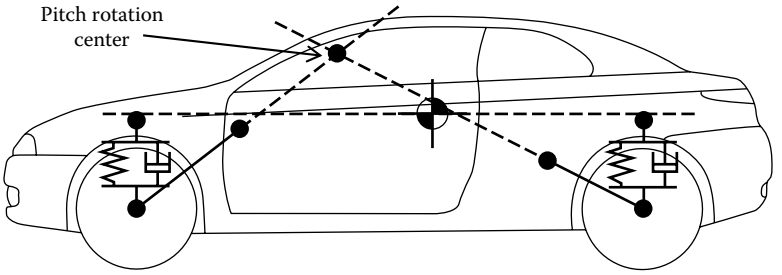


FIGURE 24.14 Front and rear suspensions arms inclinations to have a 100% anti-dive configuration (brakes on the car body).

24.4 HYDRAULIC BRAKE SYSTEM AND COMPONENTS

For cars, a hydraulic system is commonly used to transfer from the driver to the brakes both the actuation energy and the information on the amount of brake torque to be applied. Actually, the hydraulic system is rather flexible to be fitted into the car body and stiff enough to prevent undesired energy storage into the pipeline [1,3,8].

Brake by wire could be a widely adopted option in the future provided that brake mass and actuation energy will be kept as low as possible [1,8].

Figure 24.15 shows a schematic representation of a standard brake circuit. The driver force (F_P) applied at the pedal is multiplied by the lever ratio of the pedal (pedal ratio i_p) and transmitted to the master cylinder. Often, a brake booster is interposed between the pedal and the master cylinder as to increment the input force in the master cylinder. The pressure generated by the master cylinder is transmitted by a hydraulic circuit to slave pistons generating the clamping force (F_c). In this operation, the force applied to the master cylinder (F_i) is multiplied by the hydraulic circuit. In case of a circuit with an input area A_{in} at the master cylinder and an output area at the slave cylinder A_{out}, the clamping force can be computed as

$$p = \frac{F_i}{A_{in}} = \frac{F_c}{A_{out}} = const \Rightarrow F_c = \frac{A_{out}}{A_{in}} F_i = i_H F_i \tag{24.51}$$

where p is the pressure of the fluid, constant through the hydraulic circuit, and $A_{out} > A_{in}$, $F_c > F_i$, and $i_H > 1$. i_H is called hydraulic ratio and multiplies the input force. The maximum value of i_H is limited by the condition that the master cylinder must deliver the fluid needed to move the brake actuators (both for closing the clearances and for applying the force) plus the fluid absorbed by the deformation of the hydraulic circuit and maintain a residual travel. The fluid delivered by the master cylinder is equal to its volume, which can be computed as A_{in} times the travel of the master cylinder piston.

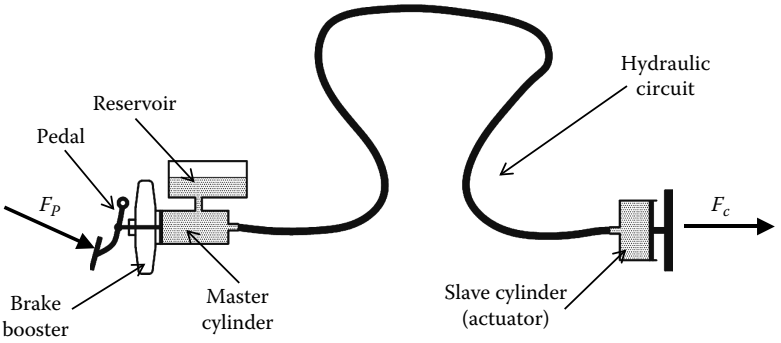


FIGURE 24.15 Schematic representation of a hydraulic brake circuit.

The travel of the piston is limited by the travel of the pedal and the pedal ratio (which is defined by the maximum force equal to 500 N for vehicles up to eight passengers and 700 N for more passengers [14]); thus, the minimum value of A_{in} is limited by the amount of fluid required. Incrementing A_{out} leads to an increment on the required fluid volume and thus is limited by A_{in}. To maximize i_H, all the component of the hydraulic circuit have to be very stiff in order to reduce the amount of fluid absorbed by their deformation. Also, the clearance between pad and disk (brake disk) or shoe and drum (drum disk) must be reduced at the minimum as to reduce the actuators' travel needed to close the clearance.

Considering the pedal ratio i_p and a force increment i_b at the booster, the total multiplication from the applied force F_P to the clamping force F_c is

$$F_c = i_P \cdot i_b \cdot i_H \cdot F_P = i_{out} F_P \Rightarrow i_{out} = i_P \cdot i_b \cdot i_H = \frac{F_c}{F_P} \tag{24.52}$$

i_{out} is called external brake system gain, defined as the ratio between the clamping force and the pedal force. The overall brake system gain i_{og}, defined as the ratio between the brake circumferential force (F_B) and the pedal force, can be obtained by multiplying the external gain by the internal gain C^* typical of each brake type:

$$i_{og} = i_{out} \cdot C^* \tag{24.53}$$

The internal gains C^* can be computed for each brake type (disk brake, drum brake) and are summarized in ISO 611 [17] for the different constructions.
For a disk brake with two friction surfaces, the internal gain can be computed as

$$C^* = \frac{F_B}{F_c} = \frac{F_{B,sch1} + F_{B,sch2}}{F_c} = \frac{F_c \mu_{lining} + F_c \mu_{lining}}{F} = 2\mu_{lining} \tag{24.54}$$

where

μ_{lining} is the friction coefficient between pad lining and disk
$F_{B,sch1}$ and $F_{B,sch2}$ are the tangential forces acting on the two surfaces of the disk

A typical value of C^* for a brake disk is 0.8 [17]. For drum brakes, the value of C^* depends on the type of construction and ranges between 1 and 5.5 [1,17,18] as reported in Table 24.1. Construction types with high values of C^* should not be employed as service brakes in order to avoid left-to-right and front-to-rear brake unbalances. These constructions are employed mostly as parking brakes.

By considering the efficiencies of the force–pressure transformation in the master cylinder (η_{act}) and of the pressure–force transformation in the slave cylinder (η_B), the brake friction radius (r_B), and the tire rolling radius (r_t), the following formula can be derived that relates the braking force (F_x) to the pedal force:

$$F_x = \underbrace{i_P \cdot i_b \cdot \eta_{act}}_{\substack{\text{Actuation} \\ \text{mechanism}}} \cdot \underbrace{i_H \cdot \eta_B}_{\substack{\text{Hydraulic} \\ \text{system}}} \cdot \underbrace{C^* \cdot \frac{r_B}{r_t}}_{\substack{\text{Brake} \\ \text{mechanism}}} \cdot F_P \tag{24.55}$$

Equation 24.55 is very approximate and can be used only under restrictive conditions, but it gives an overview of the effects of the three subsystems composing the brake system.

TABLE 24.1
Internal Gain C^* for Various Types of Brakes

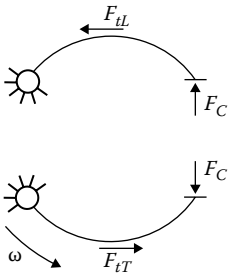
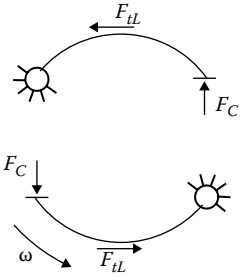
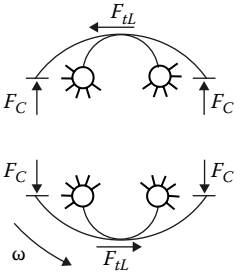
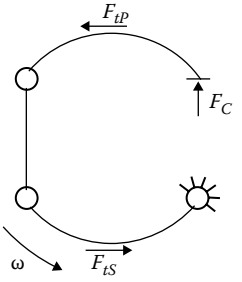
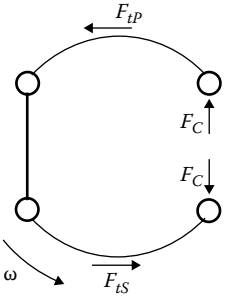
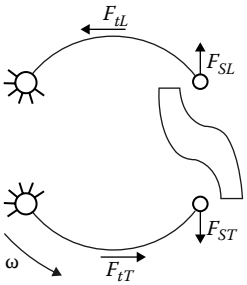
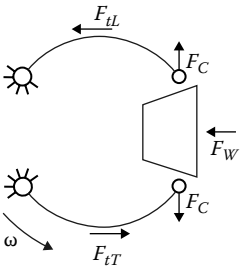
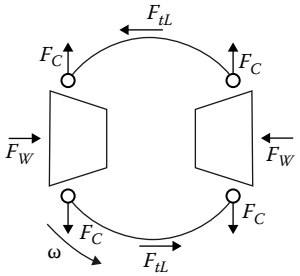
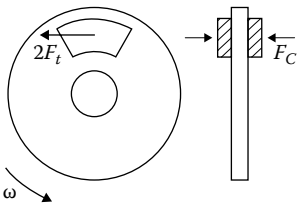
Brake Type	Force Scheme	Typical Friction Coefficient μ (from [17])	Internal Gain C^*
Simplex		0.44	$C^* = \frac{F_{tL} + F_{tT}}{F_C}$ Typical values Forward: $C^* = 2.2$ Reverse: $C^* = 2.2$
Duplex		0.40	$C^* = \frac{2F_{tL}}{F_C}$ Typical values Forward: $C^* = 3.4$ Reverse: $C^* = 1.0$
Duo duplex		0.40	$C^* = \frac{2F_{tL}}{F_C}$ Typical values Forward: $C^* = 3.4$ Reverse: $C^* = 3.4$
Uni servo		0.40	$C^* = \frac{F_{tP} + F_{tS}}{F_C}$ Typical values Forward: $C^* = 5.5$ Reverse: $C^* = 1.0$
Duo servo		0.40	$C^* = \frac{F_{tP} + F_{tS}}{F_C}$ Typical values Forward: $C^* = 5.5$ Reverse: $C^* = 5.5$

TABLE 24.1 (continued)
Internal Gain C^* for Various Types of Brakes

Brake Type	Force Scheme	Typical Friction Coefficient μ (from [17])	Internal Gain C^*
Cam		0.44	$C^* = \frac{2(F_{tL} + F_{tT})}{F_{SL} + F_{ST}}$ $F_{SL} \neq F_{ST}$ Typical values Forward: $C^* = 1.9$ Reverse: $C^* = 1.9$
Simplex wedge		0.35	$C^* = \frac{F_{tL} + F_{tT}}{F_C}$ Typical values Forward: $C^* = 2.0$ Reverse: $C^* = 2.0$
Duo wedge		0.35	$C^* = \frac{2F_{tL}}{F_C}$ Typical values Forward: $C^* = 3.4$ Reverse: $C^* = 3.4$
Disk		0.40	$C^* = \frac{2F_t}{F_C}$ Typical values Forward: $C^* = 0.8$ Reverse: $C^* = 0.8$

Sources: Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, ISBN 978-0-7680-1787-8, 2008; ISO 611, *Road Vehicles—Braking of Automotive Vehicles and Their Trailers—Vocabulary*, 4th edn., April 15, 2003; Heisle, H., *Advanced Vehicle Technology*, Butterworth-Heinemann, Oxford, U.K., 2002; Limpert, R., *Brake Design and Safety*, 2nd edn., SAE, Warrendale, PA, 1999.

F_C, *clamping force;* F_t, *tangential force (L:* leading shoe, *T*: trailing shoe, *P*: primary shoe, *S*: secondary shoe); F_W, wedge force; F_S, actuating force (*L*: leading shoe, *T*: trailing shoe); ω, rotation velocity.

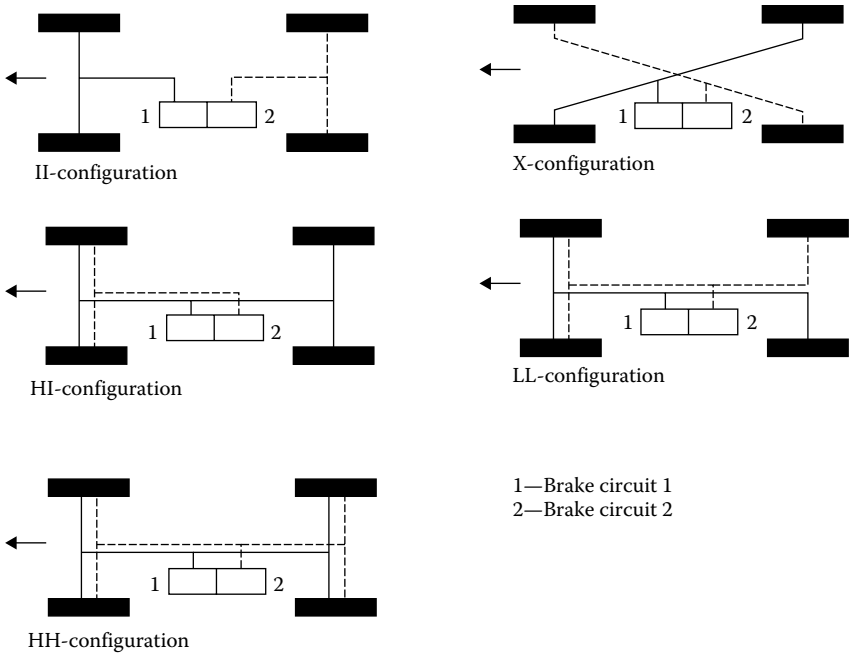


FIGURE 24.16 Brake circuit configurations according to DIN 74000. (From ISO 611, *Road Vehicles—Braking of Automotive Vehicles and Their Trailers—Vocabulary*, 4th edn., April 15, 2003; Adapted from Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, 2008, ISBN 978-0-7680-1787-8.)

24.4.1 Brake System Configurations

The legal requirements on the safety of the brake system impose the use of two separate hydraulic circuits. In case of failure of one of the circuits, a certain amount of braking power has to be still available on the vehicle. In the DIN 74000 standard [19], the five circuit configurations that can be employed on a car are described (here depicted in Figure 24.16).

Among the five possible configurations, the most employed on passenger cars are the first two (II and X configurations) [1,8]. In the last three configurations (HI, LL, and HH configurations), two different circuits reach two or more brake calipers. The calipers have to be designed in order to have two separate pistons and two inlet pipes. This construction is more expensive and thus rarely used for passenger cars and adopted only for light to heavy commercial vehicles.

Referring to the first two configurations, the II configuration is adopted for vehicles with a high load on the rear axle. In this case, if the front axle circuit has a failure, a high load is present at the rear axle and sufficient brake force is still available. By separating the front and rear circuit, rear axle brakes remain effective in the event of thermal overload on the front axle (e.g., if vapor bubbles form in a hydraulic brake). The X configuration is adopted for most vehicles with forward or none weight bias. With this arrangement, each pair of brakes diagonally opposite each other are supplied from one brake circuit. This way, if one circuit fails, the greater proportion of the braking force at the front axle is still exploited at one wheel. It is possible to compensate for the brake pull that occurs if a circuit fails as a result of the asymmetrical distribution of braking force between the left and right of the vehicle by designing the axle appropriately (e.g., with a negative scrub radius).

24.4.2 Brake System Components

The main components of the brake system are similar for all passenger vehicles. Figure 24.17 shows the scheme of a braking system (II configuration). The driver exerts a force on the pedal (1), moving the piston rod connecting the pedal to the tandem master cylinder (3) through the brake booster (2).

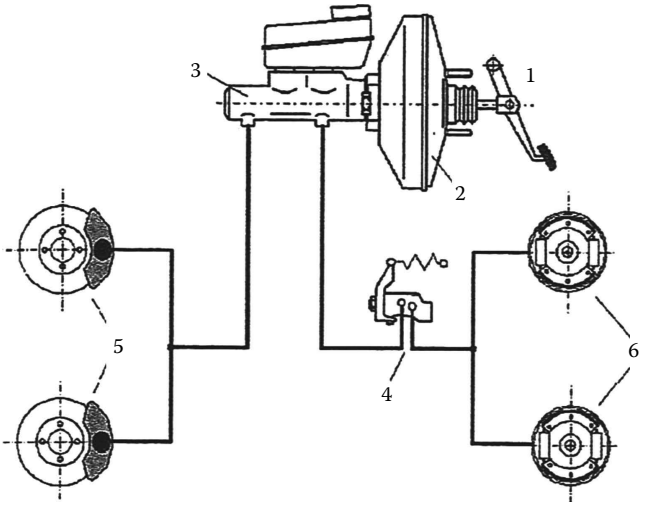


FIGURE 24.17 Brake circuit: (1) pedal, (2) vacuum brake booster, (3) tandem master cylinder with reservoir, (4) brake pressure regulator, (5) front axle wheel brakes, and (6) rear axle wheel brakes. (Adapted from Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, 2008, ISBN 978-0-7680-1787-8.)

The brake booster amplifies the force applied by the driver by introducing energy into the system. The tandem master cylinder transforms the mechanical force into pressure and feeds both of the two separate hydraulic circuits. A reservoir is connected to each chamber of the tandem master cylinder in order to provide the required brake fluid. One or two (depending on the circuits configuration) pressure regulators (4) are located on the line or lines to the rear axle brakes in order to reduce the rear pressure for high values of deceleration. Brakes (5 and 6) are located to the front and rear wheels. Disk (5) or drum (6) brakes are employed. Disk brakes can be used both at the front and rear axles, while the use of drum brakes is limited to the rear axle of low performances vehicles.

In the following sections, a brief description of each component of the brake system will be provided.

24.4.2.1 Pedal

The pedal is the interface between the braking system and the driver. The braking action can be sought of as a closed-loop control system in which the driver is the main controller (actually some control operations are demanded to the electronic control of the system, if present). In this control loop (Figure 24.18), the pedal has a double function. It is the actuator powered by the driver to

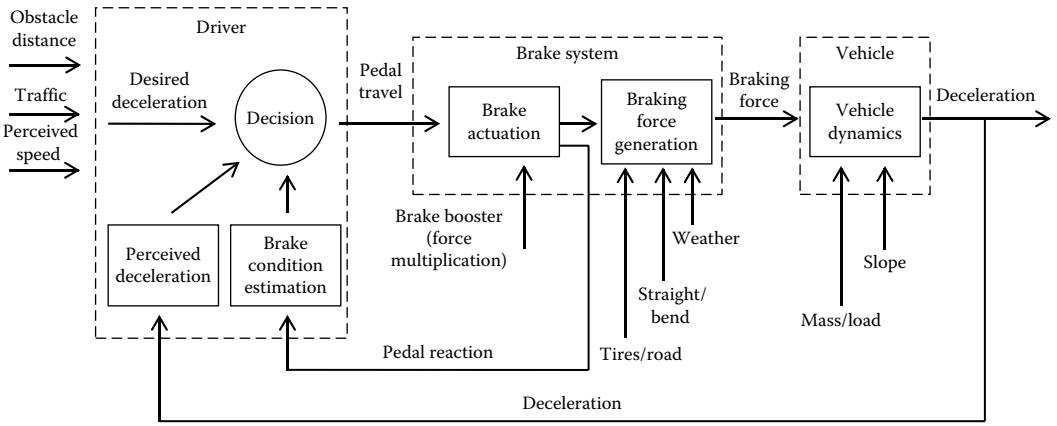


FIGURE 24.18 Brake control loop.

provide the braking energy and it is a sensor that the driver uses to estimate the state of the braking system. The ergonomics of the pedal and its force characteristics are of fundamental importance for the driver confidence and for safety. The driver force is multiplied by the pedal by a factor between 3 and 5. A correct positioning of the pedal can contribute to reduce the stopping distance, minimizing the time needed to reach the command in emergency situations.

24.4.2.2 Brake Booster

The function of the brake booster is to increment the force applied by the driver reducing the physical effort required for braking. Safety requirements prescribe that in case of booster failure, the driver must be still able to stop the vehicle by its own strength. As a consequence, very high values of boost ratio i_b cannot be employed, being difficult in this case to match this requirement. The booster has to function as a follower of the driver inputs, that is, the booster must apply its action after the driver has displaced the pedal and finish its action when the position defined by the driver is reached. Two types of brake booster are currently implemented, namely, vacuum brake booster and hydraulic brake booster [1,8,20].

24.4.2.2.1 Vacuum Brake Booster

Figure 24.19 shows the scheme of a vacuum brake booster able to comply with the requirements stated earlier. The driver action, through the pedal, is applied to the piston rod (7). The output force is transmitted to the master tandem cylinder by the push rod (1). Notice that if the system is not working, the piston rod can directly push the tandem rod allowing the driver to fulfill the braking action, even if without any boosting aid.

The operating principle is based on the difference of pressure between the vacuum chamber (2) and the working chamber (8). The vacuum chamber is connected to a void generator, either a void pump (diesel engines) or the inlet manifold (gasoline engines). A non-return valve is employed to avoid the vacuum chamber to release the void when the void generator is not functioning. When the

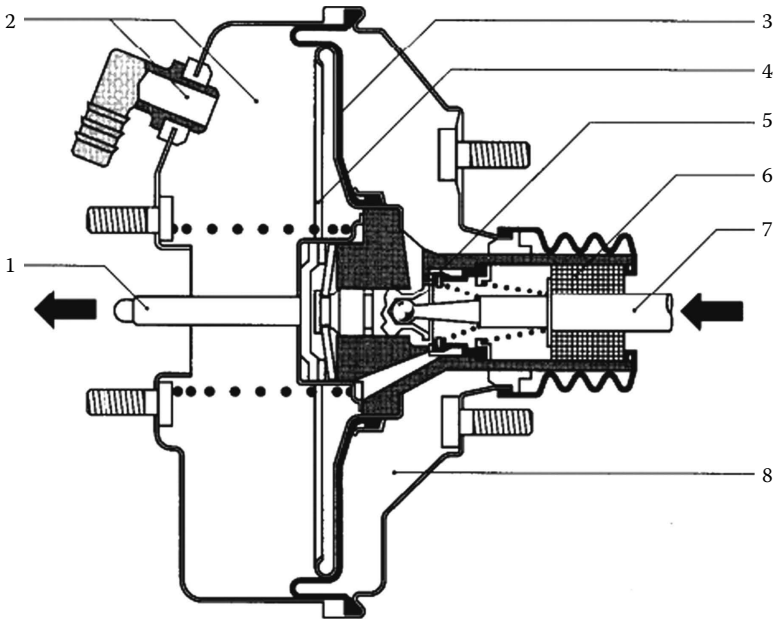


FIGURE 24.19 Vacuum brake booster: (1) push rod (output force to tandem cylinder), (2) vacuum chamber with vacuum connection, (3) diaphragm, (4) working piston, (5) double valve, (6) air filter, (7) piston rod (pedal force), and (8) working chamber. (Adapted from Bosch, *Automotive Handbook*, 7th edn., SAE, Warrendale, PA, 2007.)

system is at rest, the vacuum chamber and the working chamber are connected through the valve housing. The pressure is equal in the two chambers and no action is present. When the piston rod is displaced, the connection between the two chambers is closed, and the working chamber is connected to the atmospheric pressure by means of the double valve (5). The higher pressure in the working chamber acts on the diaphragm (3) connected to the working piston (4) that acts on the push rod adding the pressure force to the force applied by the driver, overcoming the reaction force of the returning spring. The valve housing is connected to the working piston, and when the final position is reached, the working chamber and the vacuum chamber are connected again to stop the booster action.

The active brake booster has an electrically actuated solenoid drive integrated into the housing to actuate the valve, providing support for additional functions (ESC priming, electronic brake assist, and ACC) [1].

Despite their larger dimensions, vacuum brake boosters have until now been preferred to hydraulic brake boosters. They are generally used for master cylinder volumes up to about 25 cm^3 [6], due to their limitation in booster pressure and size. The main reasons for their employment are their low manufacturing costs, a free source of vacuum at disposal in gasoline engines, and a better return feeling to the pedal [6].

24.4.2.2.2 Hydraulic Brake Booster

The hydraulic boost system uses as energy source a pressurized fluid. The brake system is a conventional brake system with the booster and plumbing added. In most cases, the source of pressurized fluid is the steering pump. A pressure-controlled flow volume regulator is employed for taking a small portion of flow supply from the power steering and charging a hydraulic pressure accumulator. The accumulator has, also, the function of providing the pressurized fluid for a certain number of brake applications in case of pump failure.

The hydraulic brake booster, compared with the vacuum brake booster, requires less installation space and it is able to furnish higher booster actions. However, it has a much worse pedal feeling and higher costs. The hydraulic brake booster is employed when a high brake boost is needed (heavy passenger cars such as limousines) [8,20].

24.4.2.3 Tandem Master Cylinder

The requirement of two distinct hydraulic circuits imposes the use of two master cylinders or a tandem master cylinder in order to pressurize both circuits. On most vehicles, the master tandem cylinder is adopted, while two master cylinders connected via a balance lever is relegated to race car applications (see Chapter 15 for details).

The tandem master cylinder must be constructed as such to provide the required pressure to the two hydraulic circuits and in case of failure of one circuit, it must be able to deliver the pressure in the other circuit. Figure 24.20 shows a typical construction for a master cylinder and its reservoirs. The reservoirs are needed in order to provide the hydraulic fluid absorbed by the displacement of the actuators, to compensate for the worn of the lining and for thermal dilatations.

The two pistons A and B are located in the same bore. Piston A is connected to the pedal, either directly or through the brake booster. The pedal displacement causes piston A to move left, and pressure is built in the chamber between the two pistons connected to the first hydraulic circuit. By effect of the pressure, a force arises on piston B that starts moving left winning the reaction force of the restoring spring and building up pressure in the last chamber connected to the second hydraulic circuit. By a proper design of the pistons areas, the same pressure can be given to the two circuits.

In case of failure of one of the two circuits, the tandem master cylinder is still able to provide pressure to the other circuit. If the circuit connected to port F fails, piston B will slide till the stop is reached without building up any pressure. As soon as the stop is reached, some reaction is present and pressure can build up in the other chamber. If the circuit connected to port G fails, piston A

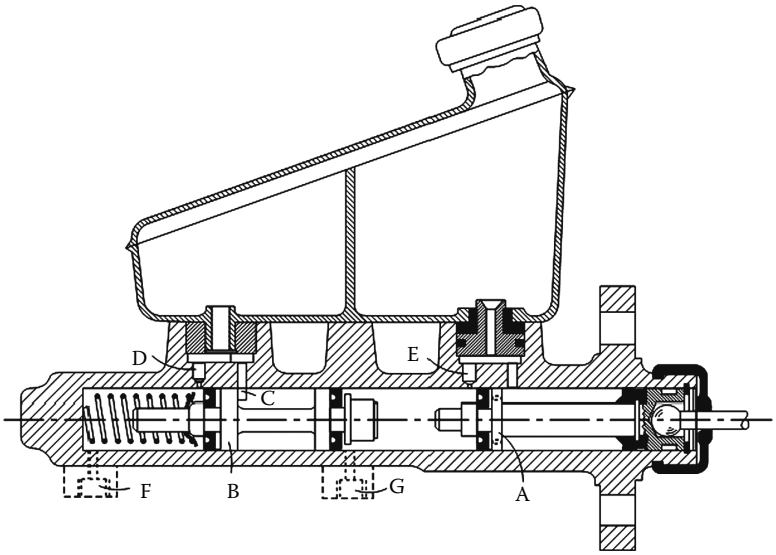


FIGURE 24.20 Master tandem cylinder. A and B, pistons; C, D, and E, reservoirs holes; F and G, connections to the two hydraulic circuits. (Adapted from Garrett, T.K. et al., *The Motor Vehicle*, Butterworth-Heinemann, Oxford, U.K., 2001.)

slides without building up pressure until it reaches piston B. At this moment, piston B is moved and pressure builds up normally in the chamber connected to port F [18,21].

24.4.2.4 Pressure Limiters and Reducers

As shown in the previous sections, the variation of the axle loads during braking is such that the braking action at the front axle increases and at the rear axle decreases. The ideal BFD is a nonlinear curve that can be conveniently approximated by limiting or reducing the brake pressure at the rear axle. Up to the mid-1990s, this operation has been realized by means of hydraulic–mechanical pressure limiters and reducers. Nowadays, these systems are increasingly replaced by electronic EBD, integrated into the ABS system [1].

In Figure 24.21, the typical characteristics of the three more diffused kinds of pressure distribution valves are shown.

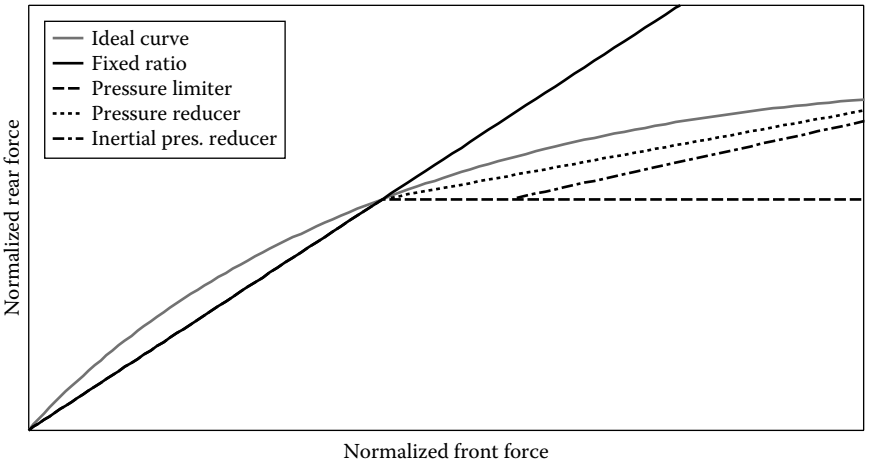


FIGURE 24.21 Characteristic of pressure limiter (dashed line), pressure reducer (dotted line), and inertial pressure reducer (dashed–dotted line).

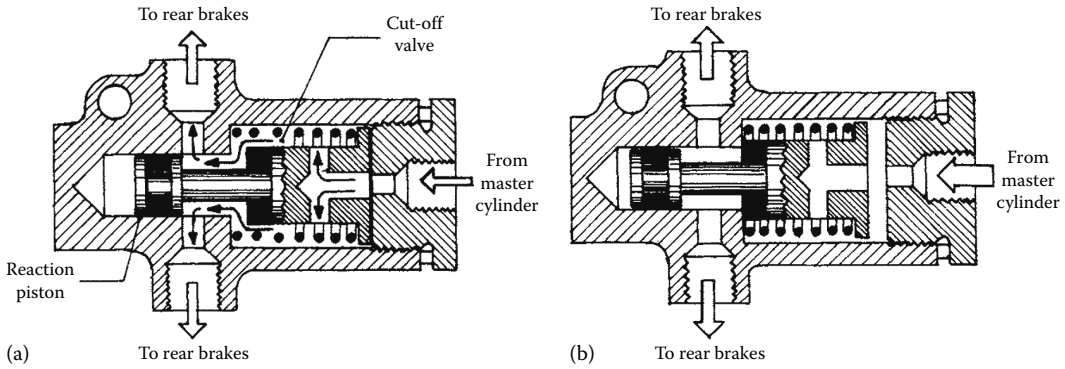


FIGURE 24.22 Pressure-limiting valve. (Adapted from Heisle, H., *Advanced Vehicle Technology*, Butterworth-Heinemann, Oxford, U.K., 2002.) (a) Low deceleration. (b) High deceleration.

In case of the pressure limiter (Figure 24.22), a cutoff pressure is defined at which the circuit of rear brakes is isolated from the master cylinder, that is, the rear brake pressure is kept constant regardless of the increase of pressure into the master cylinder. To this end, the limiting valve has an internal piston connected to a reaction spring that is moved by the pressure from the master cylinder till the manifold connecting the rear axle is closed.

Figure 24.23 shows a proportional valve reducer. In this case, the master cylinder pressure causes the reaction piston to move outward till it reaches the conical cutoff valve at the designed cutoff pressure, isolating the rear axle. Then, if the master cylinder pressure rises, the conical valve is moved inward, and the rear axle is again connected to the master cylinder. The pressure at the rear axle rises less than the pressure in the master cylinder, because as the two circuits are connected, the rise in pressure of the rear circuit makes the reaction piston to close the valve again. During high-pressure braking, this sequence is repeated and the valves continue to oscillate letting the rear pressure rise with a slower slope with respect to the master cylinder pressure. The proportional valve shown in the figure is also load sensitive. The reaction spring is connected to the suspension system of the vehicle and the preload varies as the load on the vehicle varies. In this way, the cutoff pressure can be varied and the characteristics of Figure 24.24 can be achieved [18,20,21].

In Figure 24.25 an inertial pressure reducer valve is depicted. In this case, an inertial ball is positioned on an inclined slope with respect to the direction of travel. When a certain amount of

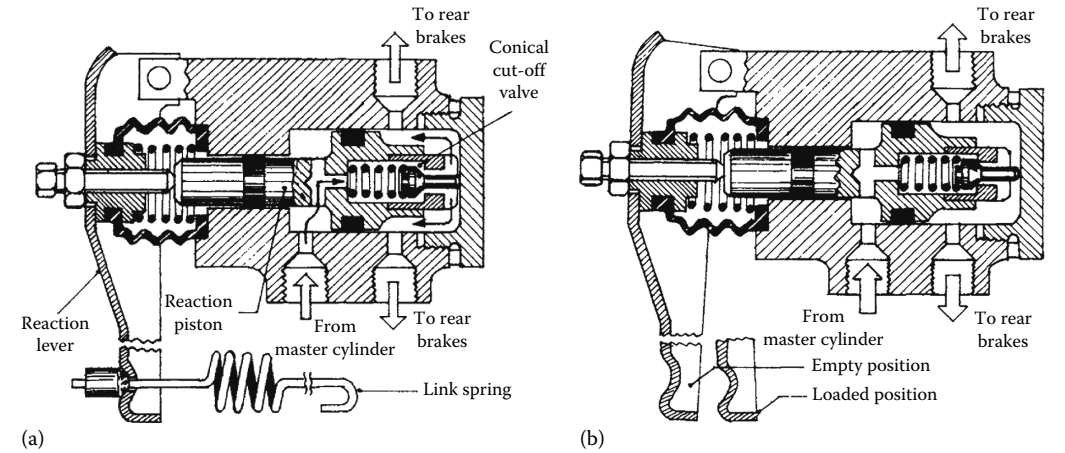


FIGURE 24.23 Load-sensitive pressure reducer valve. (Adapted from Heisle, H., *Advanced Vehicle Technology*, Butterworth-Heinemann, Oxford, U.K., 2002.) (a) Low deceleration. (b) High deceleration.

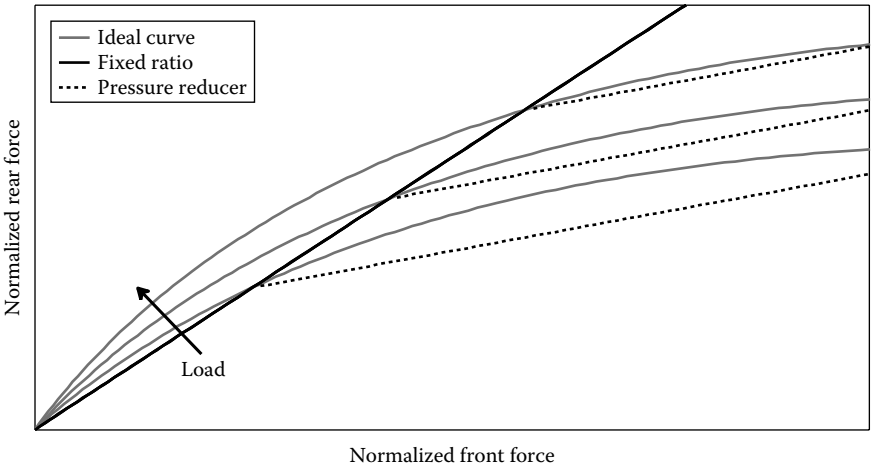


FIGURE 24.24 Load-sensitive pressure reducer valve characteristics.

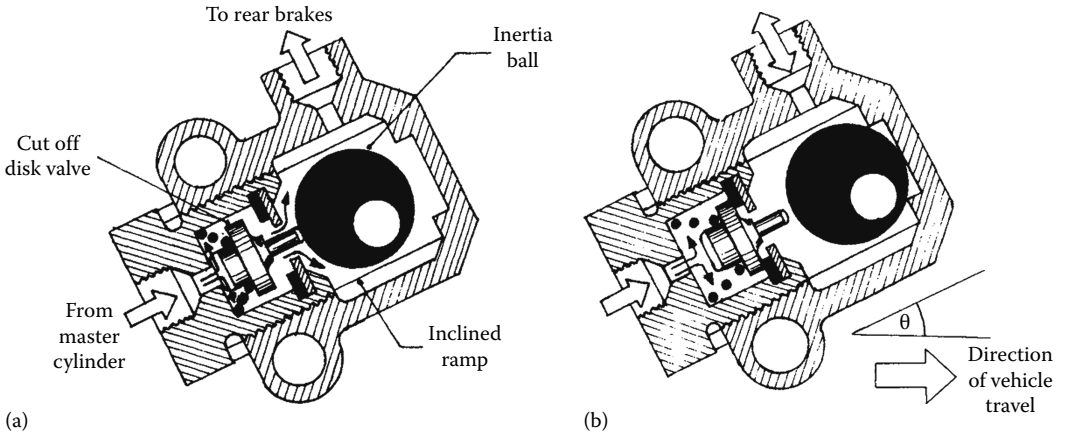


FIGURE 24.25 Inertial pressure reducer valve. (Adapted from Heisle, H., *Advanced Vehicle Technology*, Butterworth-Heinemann, Oxford, U.K., 2002.) (a) Low deceleration. (b) High deceleration.

longitudinal deceleration is reached, the ball is removed from the cutoff disk valve that closes and isolates the rear circuit. If a differential cutoff disk valve is employed, a proportional characteristic similar to the one of Figure 24.24 can be obtained [18].

24.4.3 Brake Types

For vehicle application, two different types of brakes are employed, drum brakes and disk brakes. Other types of bakes, such as band brakes, are used on tracked vehicles. Band brakes are employed in vehicle automatic transmissions with planetary gears for changing the planetary configuration and thus the transmission ratio [22] (for details on automatic transmissions, see Section 23.5.3.5).

Drum brakes have been firstly utilized on hydraulic actuated brake systems and have been the sole kind of brake in use until mid-fifties, when the disk brakes with part contact between disk and pads were employed by Jaguar [1] firstly in competition and then on mass production. Since then, disk brakes have become the common choice, while drum brakes have been relegated to the rear axle of low performances vehicles.

24.4.3.1 Drum Brakes

Drum brakes are realized by means of a drum located on the wheel hub and rotating with it. A pair of shoes, fixed with respect to the vehicle, is located inside the drum and is pressed to the drum by hydraulic actuators creating a tangential force on the drum. Different arrangements of the shoes are possible and shown in Table 24.1. Depending on shoe arrangements, the internal gain of the drum brake can be very high. In some arrangements, different internal gains are obtained for forward and reverse rotations [1,6,17,18].

Even if drum brakes have internal gains much higher than the disk brakes, their performances are poorer. This is due to many factors. Drum brakes have to be located inside the hub; thus, small brake diameters can be realized reducing the brake torque. For the same reason, the heat dissipation of these brakes is limited. Moreover, they greatly suffer thermal expansion that increases the internal diameter of the drum reducing the pressure between drum and shoes. The friction coefficient is not constant and, given the high internal gain, brake balancing problems can arise. A relatively high clearance is needed between shoes and drum, thus augmenting the volume of fluid needed to actuate the system and, consequently, reducing the hydraulic gain. Finally, drum brakes are prone to brake squeal.

Due to low manufacturing costs, drum brakes are employed on the rear axle of low-performance vehicles where thermal problems are much less relevant. Thanks to their high internal gain, they are also employed as parking brakes, even on heavy vehicles [1,6,18,21].

24.4.3.2 Disk Brakes

Disk brakes are composed of a rotating disk, called rotor, connected to the wheel hub and a couple of pads connected to a caliper attached to the stub axle. By means of hydraulic actuators, the pads clamp the disk generating a tangential force.

The caliper can be either fixed or floating. The fixed caliper (Figure 24.26a) has a fixed housing and the brake cylinders are located on both sides of the disk. This arrangement is highly stiff, ensuring low fluid displacements when operated. Since brake cylinder are located on both sides, large design space is required, and there are most employed on heavy passenger cars with rear drive with positive scrub radius.

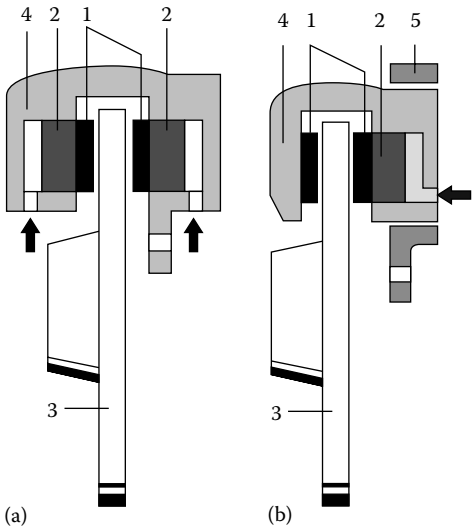


FIGURE 24.26 Disk brakes. (a) Fixed caliper and (b) floating caliper: (1) friction pads, (2) brake piston, (3) brake disk, (4) caliper, and (5) support. (Adapted from Bosch, *Automotive Handbook*, 7th edn., SAE, Warrendale, PA, 2007.)

In the floating caliper (Figure 24.26b), one or two hydraulic actuators are located at one side of the disk and the caliper is free to move axially. When the actuation force is applied, the reaction on the disk on the side of the actuators moves axially the caliper allowing for the contact between the other pad and the disk. Floating calipers require less design space and can be useful to locate properly the steering axis.

The clearance of disk brake is very small and it is guaranteed by the sealing ring of the piston that retracts the piston of a certain amount of space when the braking pressure is removed (rollback).

During braking, most of the released heat is absorbed by the disk, which can reach very high temperatures (up to 700°C). Most thermal problems of the disk brakes are related to the reduction of friction coefficient of the linings that can release vapor at high temperature (fading), to the damaging of localized zones of the disk due to microstructural modification (hot spots), to the development of bubbles in the fluid (especially for fixed calipers), and to the deformation of the rotor bell that inclines the disk with respect to the pads (dishing). To reduce disk temperature, ventilated disk can be employed [1,6,9,18,21,23].

A very important issue regarding brake disk is the vibration of the disk while braking. This problem can cause physically perceived low-frequency vibration or high-frequency vibration perceived as noise. This argument will be thoroughly explained in the last sections of this chapter.

24.5 RETARDERS

The speed regulation of heavy commercial vehicles driving downhill requires the dissipation of the potential energy of the vehicle. As shown in Equation 24.4, the thermal power that must be evacuated from the brake system is a linear function of the velocity and can be of the order of several hundreds of kilowatts. The friction brakes employed as service brakes are not able to dissipate this amount of thermal power for the time needed to run down the slope. In this situation, retarders (or continuous braking systems) are required.

Retarders are conceived in order to allow the driver to continuously regulate the speed of the vehicle when driving downhill without suffering any damage from thermal dissipation.

Three different types of retarders can be employed, namely, engine brake, hydrodynamic retarder, and electrodynamic retarder. The engine brake is also called primary retarder as it is applied to the drive input shaft, while the other two are called secondary retarders as they are applied to the drive output shaft.

Internal combustion engines, when the throttle is closed, are driven by the transmission and absorb some energy. To increase the negative torque of the dragged engine, it is sufficient to add a throttle in the exhaust pipe, as close as possible to the exhaust manifold. When required, the throttle can close the exhaust manifold; the exhaust gas remains in the exhaust pipe and during the fourth cycle of the engine (exhaust stroke), the piston has to compress the gas. After the compression, the compressed air is trapped in the exhaust pipe by the closure of the exhaust valve and cannot do positive work in the expansion cycle of the engine. In case, by adding a decompression valve, also the second cycle of the engine (compression stroke) can be used to dissipate energy. By this simple device, the engine is able to absorb up to 80%–90% of the engine power, depending on the engine speed [1,3,5,6,18].

The hydrodynamic retarder heats a fluid to dissipate the kinetic energy of the vehicle. A hydrodynamic retarder has a design similar to the one of a fluid coupling. A rotor (which has the function of a pump in the fluid coupling) is connected to a drive shaft, while a stator (which has the function of the turbine) is connected to the retarder housing. The rotor, driven by the drive shaft, applies a torque to the stator as a function of the speed of the vehicle. The stator cannot rotate and all the energy is dissipated and transformed into heat. The heat is then evacuated through the cooling system of the engine. By regulating the level of the fluid in the rotor–stator system, it is possible to regulate the torque of the retarder. Up to the 200% of the nominal engine power can be dissipated by such retarders. However, for a long working period of the retarder, the cooling system performance

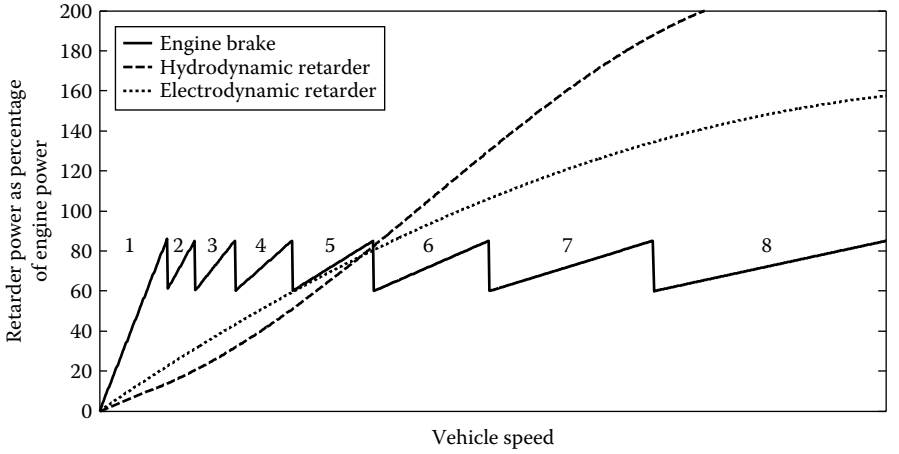


FIGURE 24.27 Retarders' braking power as a function of vehicle speed. Numbers refer to selected gear.

could be not sufficient and the retarder action has to be reduced. The advantages of such retarders are that it is impossible to lock the wheels as the torque vanishes when wheels approach locking, the retarder has virtually no wear, the load on the transmission is reduced, and the operation is less noisy than that of engine brake [1,3,5,6,18].

Electrodynamic retarders make use of eddy currents to dissipate energy. A disk, with excitation coils, is used as a stator. The rotor, connected to a drive shaft, is composed of two disks located at each side of the stator. When the coils are fed with current, a magnetic field is generated. The rotor, by rotating in the magnetic field, generates eddy currents in the coils of the stator. The current is then turned into heat and then dissipated. The amount of braking power depends on the excitation current of the stator and on the speed of the rotor [1,3,5,6,18].

Figure 24.27 shows the typical braking power of the three types of retarders as a function of speed. Notice that the engine brake, as it is applied to the input drive shaft, has a braking power depending on the chosen gear. This allows for good performances at lower speed and it is widely employed in public-transport buses. The secondary retarders have good performances at high speed, and they are mostly employed on long-distances truck for adjusting speed during high-speed downhill traveling.

24.6 NOISE VIBRATION HARSHNESS

The recurring studies on disk brake vibration and/or noise indicate that such a problem is still an elusive issue. Modeling brake vibration and/or noise involves many design variables; additionally, many harsh operational and environmental conditions have to be taken into account. Within such uncertain conditions, brake vibration and/or noise may be generated and occur. Understanding the contact mechanisms, the tribology, the nonlinear dynamics, the interactions between the brake system and the suspension system, and so forth is essential.

Even if a large amount of research has been performed and the main sources have been understood; a final solution for brake squealing seems to be far away. The only way to validate a noiseless brake system is an integrated process with a mixture of both experimental and analytical activities.

The only analytical approach available at the moment for an industrial application is the complex eigenvalue extraction in a full FEM braking system model. It seems to be (although the theoretical limitations), for practical issues, the only possibility for a robust design investigation. Brake squeal solutions in fact, can be better evaluated through a robust design approach.

Brembo has also adopted an FEM model and the complex eigenvalue analysis. This tool (conveniently tuned using experimental tests) is able to foresee the most critical situations and the tendencies.

The main goal of such a simulation is the "global" design of the components of the system (stiffness, mass distribution, modal participation, and so on) and should be used in a very early stage of the design and product development.

Unfortunately, a robust solution would frequently require that the whole suspension and braking system design would be oriented to this, especially for low frequencies squeals (1500–2500 Hz). This is not always possible: Suspensions and knuckles are designed for different targets and brake systems have to fulfill several constraints (weight, fluid consumption, thermal loads, performances in terms of friction, and, obviously, costs).

Moreover, the behavior of the friction material in terms of noise propensity is quite unpredictable because of the spread of operational conditions during braking (pressure, temperature, wear, etc.).

The global scenario is actually quite complex.

This leads to the necessity (in addition to analytical tools) of a product development made of bench and vehicle tests. During product development, in a reliable way, some countermeasures can be tested and selected that normally do not impact on global design and performances (added masses, chamfer on friction surfaces, shims, etc.) and a fine tuning with simulation results can be reached (tools as modal acquisitions and laser vibrometry are fundamental in order to reach a high correlation).

In conclusion, an integrated approach has to be implemented: analytical tools for the first design phase to get the global design of components and a fine bench and vehicle tuning in order to fix all possible squealing occurrences.

24.6.1 NVH State-of-the-Art Reviews

Many state-of-the-art reviews on vibration and/or noise generated by brakes have been produced. From these works, it is possible to understand that avoiding brake vibration and/or noise is still a cumbersome issue. Much progress has been made on understanding the brake vibration and/or noise mechanisms and causes. However, new mechanisms (unreferenced even in the recent state-of-the-art reviews) can still be found [24] or envisaged [25–27].

The following authors have written their respective state-of-the-art review, focusing on a particular point of view:

- North (seemingly the first review on brake vibration and/or noise [28])
- Nishiwaki (the first modern state-of-the-art paper on brake vibration and/or noise [29])
- Crolla and Lang (a comprehensive state-of-the-art review focusing on future issues regarding FEM [30])
- Ibrahim (two detailed state-of-the-art review focusing on disk vibration modes and nonlinear effects such as chaos [31] and [32])
- Yang and Gibson (a comprehensive review with emphasis on materials [33])
- Mottershead (a review on the dynamics of disks subjected to frictional loading, with particular emphasis on the mathematical modeling [34])
- Papinniemi et al. (a comprehensive review with emphasis on acoustics [35])
- Kinkaid et al. (a prominent state-of-the-art review on brake squeal [36])
- Ouyang et al. (a useful complement to the previous state-of-the-art review [37])
- Chen (a review from a OEM perspective [38])
- Hoffman and Gaul (a review stating well the objectives and the still open issues, with a timely focus on the future [25])
- Cantoni et al. (a comprehensive review on brake squeal modeling and testing techniques [39])

Two SAE (Society of Automotive Engineers) books [1,40] have been published focusing on brake vibration and/or noise.

In the appendix (Table 24.8), a table is presented where all of the papers cited in this chapter are sorted according to the sections in which this chapter has been subdivided.

24.6.2 Phenomena and Influencing Factors

In general, brake vibration and/or noise can be classified into many categories based on either the occurring frequencies or excitation sources (Papinniemi et al. [35], Kinkaid et al. [36], Ouyang [37], Chen [38], Dunlap [41]). We will here classify vibration and/or noise as a function of frequency. Excitations or mechanisms generating vibration and/or noise will be dealt with in the sequel.

According to [42], depending on the frequency of vibration and/or noise occurrence, there are three main categories, namely, low-frequency vibration and/or noise (0–1,000 Hz, subdivided into three subcategories), low-frequency squeal noise (1,000–3,000 Hz), and high-frequency squeal noise (3,000–15,000 Hz).

As shown in Figure 24.28, in the low frequency range (0–1000 Hz), there are two different types of structural vibrations, named judder and groan, and their associated airborne noise called moan and howl (or hum).

With respect to the huge number of publications dealing with high-frequency vibration and/or noise, (i.e., brake squeal, see Section 24.6.5.3), judder and groan received much less attention. However, this type of vibration and/or noise receives a growing interest from the automotive industry because it primarily affects the comfort (the driver experiences judder as vibrations in the steering wheel, brake pedal, and floor, see Table 24.2) and could lead, when impacting on inexperienced drivers for the first time, to faulty reactions and reduced driving safety [43–47].

The vibration frequency of the brake judder (0–100 Hz) is proportional to the wheel speed and therefore to the velocity of the vehicle. It is common to relate the judder frequencies to the wheel speed (e.g., frequencies at twice the number of wheel rotations per second are called second-order judder) [45].

Table 24.2 summarizes the judder phenomenon referring to excitation sources, transfer system, and effects.

Groan (often called creep groan [48]) generally occurs at low vehicle speed and occurs with cold brakes and high humidity (Figure 24.29). Figure 24.30 summarizes the factors that contribute to generate a creep groaning brake.

From literature [42], the squeal noise is classified as a function of frequency. There is a *low-frequency squeal* and a *high-frequency squeal* (see Figure 24.28). Low-frequency squeal is generally classified as a noise having a narrow frequency bandwidth in the frequency range above 1000 Hz, while high-frequency brake squeal is typically classified as a noise occurring at frequencies above 3 kHz. Since it is a range of frequency that affects a region of high sensitivity in the human ear, high-frequency brake squeal is considered the most annoying type of noise. The so-called wire brush noise occurs above 1 kHz.

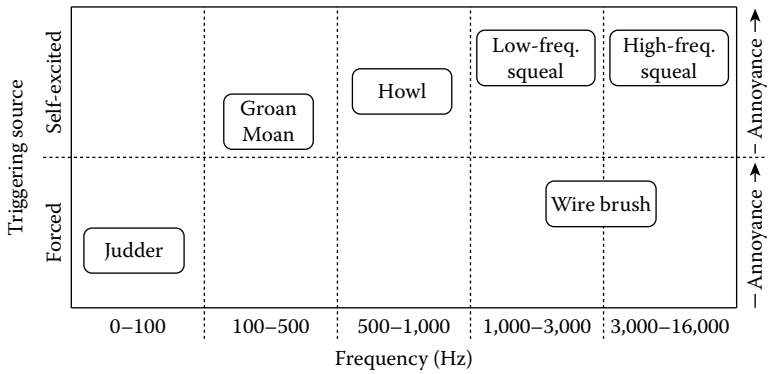


FIGURE 24.28 Brake noise classification based on the frequency of occurrence and excitation source. (From Dai, Y. and Lim, T.C., *Appl. Acoust.*, 69, 196, 2008.)

TABLE 24.2
Excitation Sources, Development, and Effects of Brake Judder

Brake Judder		
Excitation Sources	**Transfer System**	**Effects**
Wheel brake (brake pressure and/or brake torque variations) *Road spectrum* *Unbalance of rotating components* (wheel, wheel hub, brake disk) *Bearing design and clearance*	**Braking System** **Steering System** **Wheel Suspension**	Brake pedal pulsation Vibration of wheels Steering wheel rotational vibrations Body vibrations Seat vibration Low-frequency noises

Sources: Adapted from Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, 2008, ISBN 978-0-7680-1787-8; Jacobsson, H., Aspects of disc brake judder, *Proc. Inst. Mech. Eng., Part D*, 217, 419, 2003.

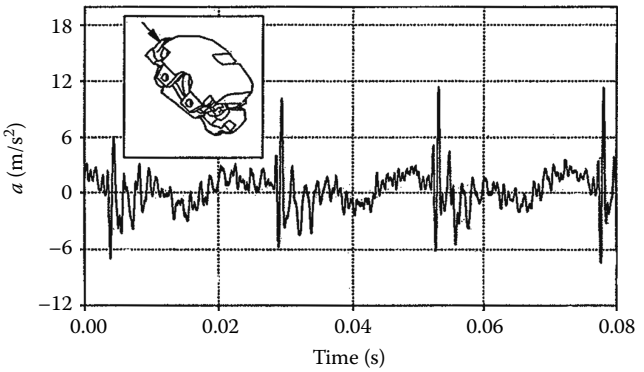


FIGURE 24.29 Tangential caliper acceleration during creep groan noise. (From Brecht, J. et al., Mechanisms of brake creep groan, Technical Report, SAE, Paper No. 973026, Warrendale, PA, 1997.)

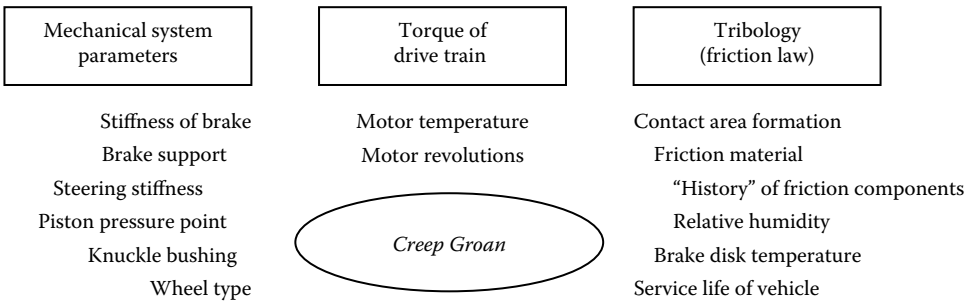


FIGURE 24.30 Factors influencing creep groan phenomenon. (Adapted from Brecht, J. et al., Mechanisms of brake creep groan, Technical Report, SAE, Paper No. 973026, Warrendale, PA, 1997.)

Nevertheless, there is no a unique idea about the frequency that divides the first phenomenon from the second one. For this reason, many authors [49,50] consider the number of the nodal diameters (which characterize the vibration modes of the disk) as a way of classification for the brake squeal noise.

In addition, the most significant difficulty in brake research is the fugitive nature of brake squeal (it can sometimes be non-repeatable) [36,51]. A brake system may not always squeal given the "same"

conditions [40]. Alternatively, small variations in operating temperature, brake pressure, rotor velocity, or coefficient of friction may result in differing squeal propensities or frequencies [38].

24.6.3 Experimental and Computational Activities

24.6.3.1 Experimental Activities

24.6.3.1.1 Displacement/Vibration

Brake tests represent a crucial method to characterize the vibrational behavior of a brake system and of its components in order to provide a ready solution for vibration and/or noise problems. However, it should be noted that the squeals cannot be reproduced or cannot be reproduced at all working conditions.

A number of different test procedures (often proprietary of OEM or brake manufacturers) exist to assess the vibration and/or noise propensity of a brake system. In such procedures, it is difficult to include all conditions under which a brake vibration and/or noise might occur due to the constraints of time and cost (e.g., it is difficult to reach some conditions of both cold and high humidity [38]).

A preliminary and fundamental analysis to detect brake vibration is just measuring displacements/velocities of the brake components. Such displacements are often very small and are of the same order of magnitude as the surface irregularities of the disk [52]. Such analyses begun few decades ago and are continuously updated. In Table 24.3, the measured displacements of a squealing disk are reported.

Optical techniques for deriving addressed displacements include

- Double-pulse holographic interferometry (DPHI) and speckle interferometry. In [36,53], many brief reports are given. These methods are used to examine the modes of squealing brakes. DPHI dominated the literature until the use of another optical technique (electronic speckle pattern interferometry [ESPI]) appeared.
- ESPI and laser vibrometer became popular in the late 1990s.
- Scanning laser vibrometer (SLV) is used to identify locations with large amplitude and ideal to identify hot spots (for more details, see [53]).

TABLE 24.3
Measured Displacements in a Vibrating Brake Disk

	Displacement Value (μm)		
Displacement Component	**Maximum**	**Minimum**	**Mean**
Upper disk section			
d_x	3.527	–4.845	–0.727
d_y	5.701	–4.139	–0.202
d_{xy} (in-plane)	6.704	—	1.743
d_z (out-of-plane)	3.106	–2.375	–0.022
Lower disk section			
d_x	4.515	–7.306	–0.306
d_y	4.437	–6.409	–0.202
d_{xy} (in-plane)	8.932	—	1.536
d_z (out-of-plane)	3.964	–3.934	–0.136

Source: Fieldhouse, J.D. et al., *Proc. IMechE*, *Part D*, 222, 1121, 2008.

The recently developed laser vibrometers can measure 3-D velocity distributions with three laser vibrometer heads. The system is able to capture the out-of-plane motion during squeal inducing from any in-plane vibration. Similar to 3-D laser vibrometer, a 3-D pulsed-laser ESPI also uses three cameras to deduce 3-D.

Another approach is the combination of accelerometer with a laser technique. Small accelerometer can be instrumented inside rotor fins in three directions as reported in [38,51].

In the literature devoted to experimental studies of brake vibration and/or noise, few details are provided on the actual brake assembly under test. As a result, given the wide range of designs and material selections that are possible, it is almost impossible to compare the results derived from different research groups. Together with Kinkaid [36], we summarize the conclusions reached in the experimental literature on brake vibration and/or noise. A braking system can squeal at a number of distinct frequencies with amplitude of the order of microns (actual deformations due to sustained brake force can reach the order of 0.1 mm); standing or traveling wave can run along the disk, rotating speed, and brake pressure and coefficient of friction of the pad can influence brake vibration and/or noise (see Figure 24.31).

The definition of the parameter values to be introduced in the mathematical models is a typical experimental activity. A basic method for friction measurement is reported in [54]. Such a method implies friction to be a parameter and is more suited for tribological scientific studies than applications involving dynamical system models.

The orthotropic characteristic of the pad liner (Young's modulus higher in the vertical than in the horizontal—contact surface—direction) can be measured by means of nonstandard material testing techniques ([55], Chapter 12); five independent elastic constants (Equation 24.56) must be measured in order to completely determine the (linear) mechanical properties. Ultrasound techniques have been adopted successfully by D. Yuhas and M. Yuhas to measure the elastic constants that are reported in Figure 24.31.

$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xz} \\ \tau_{yz} \\ \tau_{xy} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(c_{11} - c_{12}) \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xz} \\ \gamma_{yz} \\ \gamma_{xy} \end{bmatrix} \tag{24.56}$$

Temperature has an important effect on stiffness parameters of Figure 24.32; a reduction ranging from 20% to 35% can be expected raising the temperature from 20°C to 200°C [40].

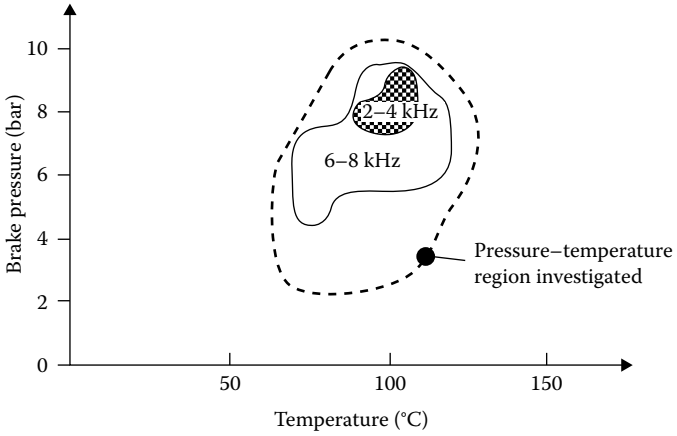


FIGURE 24.31 An example of the interaction of brake pressure and temperature on brake squeal. (Adapted from Bergman et al., *Tribotest J.*, 5(3), 265, 1999.)

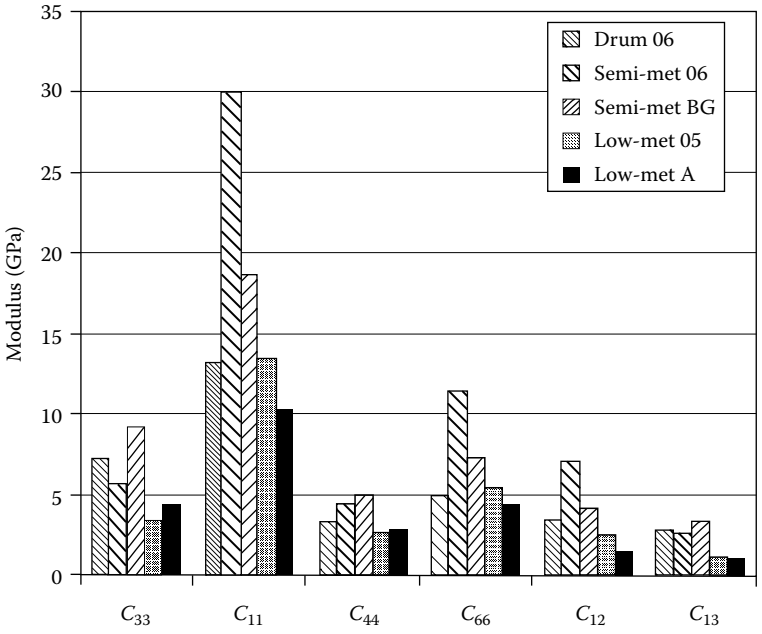


FIGURE 24.32 Comparison of moduli of friction material. (Adapted from Chen, F. et al., *Disc Brake Squeal—Mechanism, Analysis, Evaluation, and Reduction/Prevention*, SAE International, Warrendale, PA, December 2005, Chapter 12, ISBN 978-0-7680-1248-4.) The values c_{ij} $(i, j = 1, 6)$ refer to the components of the stiffness matrix in Equation 24.56 where $\sigma_x\sigma_x\sigma_x\tau_{xy}\tau_{yx}\tau_{xz}$ are the stresses and $\varepsilon_x\varepsilon_x\varepsilon_x\gamma_{xy}\gamma_{yx}\gamma_{xz}$ are the deformations.

A recent SAE test procedure (J2598, [56]) on pad damping measurement provides a robust procedure with a certain degree of reproducibility. It proposes using a small impact hammer, in order to avoid attachments like shakes (although the effect of the contact cannot be totally eliminated), and a small accelerometer. An alternative way to get test results with consistent repeatability is to use noncontact excitation (magnetic force, one of the first researchers that used it were Nishiwaki et al. [57]) and noncontact signal pickup (laser vibrometer).

24.6.3.1.2 Friction

Bergman, Eriksson, Jacobson, and coworkers correlated changes in μ_d (coefficient of friction) to the number of brakings and squeal generation. Their results are summarized in [58–67] and are consistent with what is reported in [26]. They noted that μ_d gradually increased from a value just above 0.2 and, after over 1000 brakings, began to plateau at a value below 0.6. As μ_d increased beyond a critical value, they observed that squealing behavior sharply increased (see Figure 24.33).

In [36], experimental results referring to the effects of various solid lubricants, influence of humidity, and pad morphology on brake squeal generation is reported.

24.6.3.2 Computational Activities

Computational activities are crucial for understanding brake vibration and/or noise. Simulation and analysis methods may be divided into two big categories: *complex eigenvalue analysis* in the frequency domain and *transient analysis* in the time domain [37]. The transient results can be converted to the frequency domain by a Fourier transform, provided that the phenomenon is periodic, which is not always the case.

The complex eigenvalue analysis studies the brake vibration and/or noise phenomenon at quasi-static steady sliding conditions. The equations of motion derived from a mechanical model are linearized.

When the effects of nonlinearities are present and effective, only a time-domain simulation is the proper way to analyze the dynamic behavior of the brake system. Table 24.4 summarizes the main aspects (positive and negative) of these two approaches that were stated and dealt with in [37,68–73].

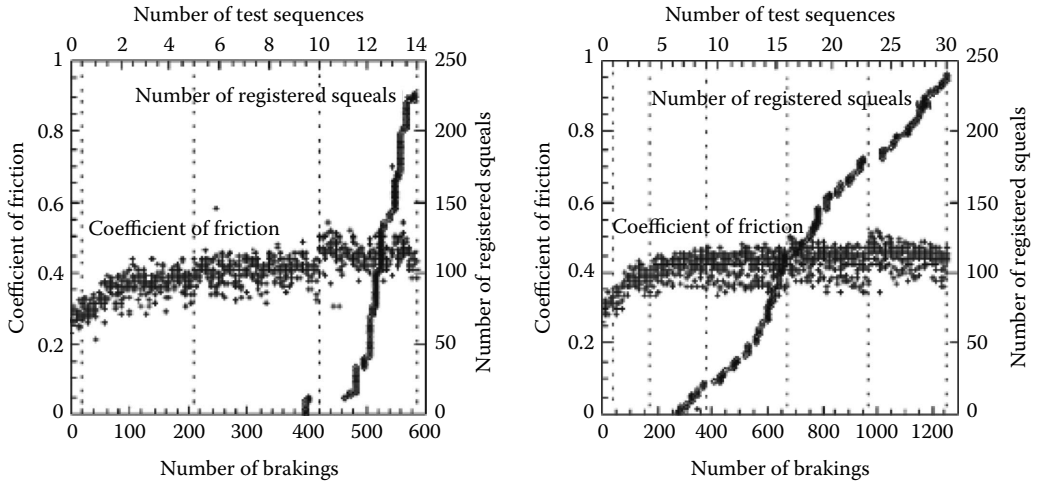


FIGURE 24.33 Cumulated occurrence of squeals and coefficient of friction μ_d as a function of the number of brakings. (From Bergman, F., Tribological nature of squealing disc brakes, PhD thesis, Uppsala University, Uppsala, Sweden, 1999; Eriksson, M. and Jacobson, S., *Tribol. Int.*, 33, 817, 2000.)

24.6.3.3 Both Experimental and Computational Activities

Since late 1930 until 1990, few contributions have been really effective in matching computation data and experimental results. In the last two decades, the ever-increasing computation power has allowed researchers to face the simulation of the complex problem of brake vibration and/or noise.

The appendix (Table 24.8) reports a list of papers dealing with both experiments and computer simulations.

A comprehensive study [74] aims at improving complex and simple brake models by means of experimental activities. In particular, it is found (by an FEM 3-D thermomechanical simulation together with a detailed analysis of experimental results) that the thermoelastic "buckling instability" is a significant factor causing hot spotting and hot judder [47].

A number of researchers have built laboratory brake assemblies in order to control the many uncertainties and shape details that are present in an actual brake and that are difficult to be modeled. In [75], a pin-on-disk (POD) system is used that provides desirable qualities of mechanical simplicity and controllable dynamic characteristics; the predicted unstable modes were actually unstable, but the stable modes could evolve toward instability. In [76], the interaction between a disk brake and its pads has been treated analytically and experimentally by means of a laboratory brake. Transverse and circumferential disk–pad oscillations were considered under deterministic and random friction coefficient. In [77], the authors study a simple lumped mathematical model and propose, by means of simple test rig (in which a simple L-shaped beam was used), two experimental methods for the prediction of squeal noise, the results rely on accurate definition of friction and tangential/contact stiffness. In [78,79], in-depth scientific analyses were undertaken focusing on a simple brake system built to control the parameters influencing brake vibration and/or noise. The crucial role of time-domain simulation was highlighted.

24.6.4 Models

24.6.4.1 Friction

In [36], an illuminating review on the relationship friction–braking systems is presented. Besides the classical books on tribology [80–83], recent contributions on friction for control application are reviewed in [84]. From [84], it appears that the historical friction models by Amontons, Coulomb, and Stribeck should be improved by state-dependent friction laws; indeed, friction has a dynamic behavior. Phenomena such as predisplacement, rate dependence, and hysteresis have been observed

TABLE 24.4
Out-of-Plane Modes of a Brake Disk

Mode	Type m = Number of Nodal Circles n = Number of Nodal Diameters	Frequency Range (Hz)	Shape
1ND	$(m, n) = (0, 1)$	700–1,100	
2ND	$(m, n) = (0, 2)$	800–1,700	
3ND	$(m, n) = (0, 3)$	1,800–3,500	
4ND	$(m, n) = (0, 4)$	2,900–4,300	
5ND	$(m, n) = (0, 5)$	4,300–5,900	
6ND	$(m, n) = (0, 6)$	5,700–7,800	
7ND	$(m, n) = (0, 7)$	7,100–9,900	
8ND	$(m, n) = (0, 8)$	8,500–12,400	
9ND	$(m, n) = (0, 9)$	12,600–13,200	
…	…	…	
…	…	…	
1NC	$(m, n) = (1, 0)$	1,700–4,000	
2NC	$(m, n) = (2, 0)$	3,400–5,700	
…	…	…	

Sources: Shape of the vibrating brake disc adapted from Abu Bakar, A.R. et al., Modelling and simulation of disc brake contact analysis and squeal, PhD thesis, Department of Engineering, University of Liverpool, Liverpool, U.K., 2005; Abu Bakar, A.R. et al., *J. Teknol*., 45(A), 1, 2006; Abu Bakar, A.R. and Ouyang, H., *Int. J. Veh. Des*., 46(4), 416, 2008; Cho, H.-J. and Cho, C.-D., *Proc. IMechE, Part D J. Automob. Eng*., 222, 895, 2008; Giannini, O. et al., *J. Sound Vib*., 292, 1, 2006; Casini, P. et al., *Nonlinear Dyn*., 46, 259, 2006; Giannini, O. and Sestieri, A., *J. Sound Vib*., 296, 583, 2006; Massi, F. et al., *Mech. Syst. Signal Process*., 21, 2374, 2007; Giannini, O. and Massi, F., *J. Sound Vib*., 310, 394, 2008; Jearsiripongkul, T. and Hagedorn, P., *Proc. Appl. Math. Mech*., 4, 93, 2004; Kung, S.-W. et al., Complex eigenvalue analysis for reducing low frequency brake squeal, Technical Report, SAE, Paper No. 2000-01-0444, Warrendale, PA, 2000; Malosh, J.B., Disc brake noise reduction through metallurgical control of rotor resonances, Technical Report, SAE, Paper No. 982236, Warrendale, PA, 1998; Masoomi, M. et al., *Polym. Compos*., 27, 461, 2006; Matsuzaki, M. and Izumihara, T., Brake noise caused by longitudinal vibration of the disc rotor, Technical Report, SAE, Paper No. 930804, Warrendale, PA, 1993; Ouyang, H. et al., *IMechE J. Automob. Eng*., 217(10), 867, 2003; Ouyang, H. et al., *Acta Mech. Sin*., 20(4), 418, 2004; Cao, Q. et al., *Int. J. Numer. Methods Eng*., 61, 1546, 2004; Joe, Y.-G. et al., *Int. J. Automot. Technol*., 9(2), 161, 2008; Trichés Jr., M.J. et al., *ABCM J. Braz. Soc. Mech. Sci. Eng*., XXVI(3), 340, 2004; Lee, H., Modal acoustic radiation characteristics of a thick annular disk, Dissertation of The Ohio State University, Columbus, OH, 2003.

experimentally and are reproduced only by dynamic models [84]. The Dahl model [85] is a state-dependent friction law defined by

$$F = F_c\left(1 - e^{-\sigma_0 |x|/F_c}\right)\mathrm{sgn}\left(\frac{dx}{dt}\right) \tag{24.57}$$

where
F is the force
F_c is the Coulomb friction
x is the displacement
σ_0 is the stiffness

After differentiation at $x = 0$,

$$\frac{dF}{dx} = \sigma_0\left(1 - \frac{F}{F_c}\mathrm{sgn}\left(\frac{dx}{dt}\right)\right) = \sigma_0\left(1 - \frac{F}{F_c}\mathrm{sgn}(\nu)\right) \tag{24.58}$$

Introducing $z = F/\sigma_0$ as a state variable (Equation 24.59),

$$\frac{dz}{dt} = \frac{1}{\sigma_0}\frac{dF}{dx}\frac{dx}{dt} = \frac{1}{\sigma_0}\frac{dF}{dx}\nu = \nu - \frac{\sigma_0}{F_c}|\nu|z \tag{24.59}$$

At steady state,

$$F_{ss} = \sigma_0 z_0 = F_c\,\mathrm{sgn}(\nu) \tag{24.60}$$

the Coulomb friction force is found. The Dahl model does not contain Stribeck effect (i.e., the coefficient of friction decreasing with increasing velocity) but the LuGre model does [85]. The LuGre friction model (to be compared with the Dahl model introduced before) reads

$$\dot{z} = \upsilon - \sigma_0\frac{|\upsilon|}{g(\nu)}z = \upsilon - h(\upsilon)z; \quad F = \sigma_0 z + \sigma_1\dot{z} + f(\upsilon) \tag{24.61}$$

where
υ is the velocity between the surface in contact
σ_1 represents microdamping
$f(\upsilon)$ represents macrodamping (typically $\sigma_0\upsilon$)

At steady state,

$$F_{ss}(\upsilon) = g(\upsilon)\mathrm{sgn}(\upsilon) + f(\upsilon) \tag{24.62}$$

The LuGre model may be considered a passive system if and only if

$$\sigma_2 > \sigma_1\frac{F_s - F_c}{F_c} \tag{24.63}$$

More details can be found in [85] with a discussion on the nonlinear dynamics of stick-slip phenomenon (Figure 24.34).

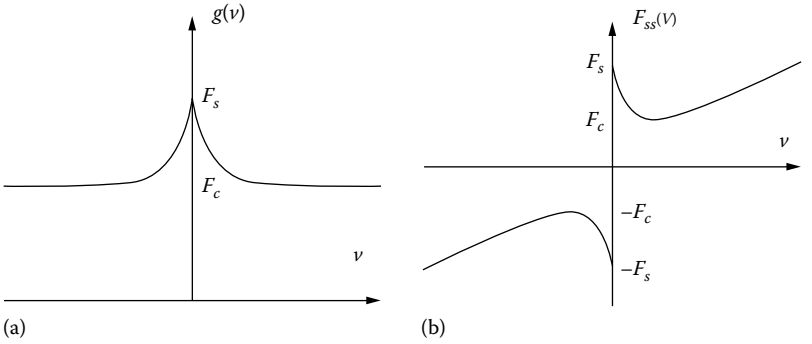


FIGURE 24.34 Functions that characterize the LuGre friction model. (a) shows the function $g(v)$ that captures Coulomb friction and the Stribeck effect, while (b) shows the steady-state friction function $F_{ss}(v)$, where $f(v)$ represents viscous friction, typically, $f(v) = \sigma_2 v$. F_c is the Coulomb friction force, whereas F_s sdenotes the stiction force. Asymmetric friction behavior can be obtained by letting $g(v)$ and $F(v)$ have different shapes for positive and negative velocities. (Adapted from Dahl, P.R., *AIAA J.*, 14, 1675, 1996.)

The measurement of friction is defined by a well-defined standard [54]. In [86], it was pointed out how difficult it is to derive the Stribeck characteristics from essentially quasi-static measurements.

From experience, the *mean* friction coefficient in the contact region between pad and disk ranges from 0.2 to 0.7, with more common values between 0.3 and 0.6. Such mean values refer to a broad contact area and are not to be expected to be smoothly distributed on the contact area. On single small contact regions, the friction coefficient measurement can be made with standard laboratory dynamometers, called tribometers. Such a measurement is interesting from the material science perspective; it has, at the moment, limited significance for common engineering applications, such as the measurement of the friction between the pad and the disk. The contact pressure ranges from 0 to 4 MPa within the contact area between pad and disk. It depends on rotor temperature that rises up to 500°C.

In [27,87], a stochastic approach was attempted to model friction brake forces, since it is evident the random nature of brake vibration noise. The contact surface and the relative speed were found to be random, leading directly to random friction force. A stochastic definition of friction was introduced as

$$\mu = \left(\mu_k + \frac{\mu_s - \mu_k}{\cosh(\alpha v_{rel})} \right) \tanh(\beta v_{rel}) \tag{24.64}$$

where

μ_s and μ_k are the static and kinetic friction coefficient, respectively
α and β are shape parameters
v_{rel} is the relative velocity
the friction coefficient is considered a random variable with a Gaussian white noise process of mean μ_k

The implementation of a stochastic equivalent (mean) friction coefficient might explain the elusive nature of brake vibration and/or noise, but the physical meaning could be lost. A much more interesting approach that captures the friction coefficient (random) variation by devoting attention to phenomena occurring in the micro- (micron) or meso- (micron millimeter) scale is due to the scientific research by Ostermeyer on the topic (see [26] where previous papers are cited). In [26], the variation of the friction force is related to the synchronized micro vibrations of the hard particles appearing on the contact surface between the pad and the disk. An example of a mechanical model is shown in Figure 24.35 where the hard particles are elastically connected both to the polymeric substrate and to each other. A macroscopically varying friction can be computed.

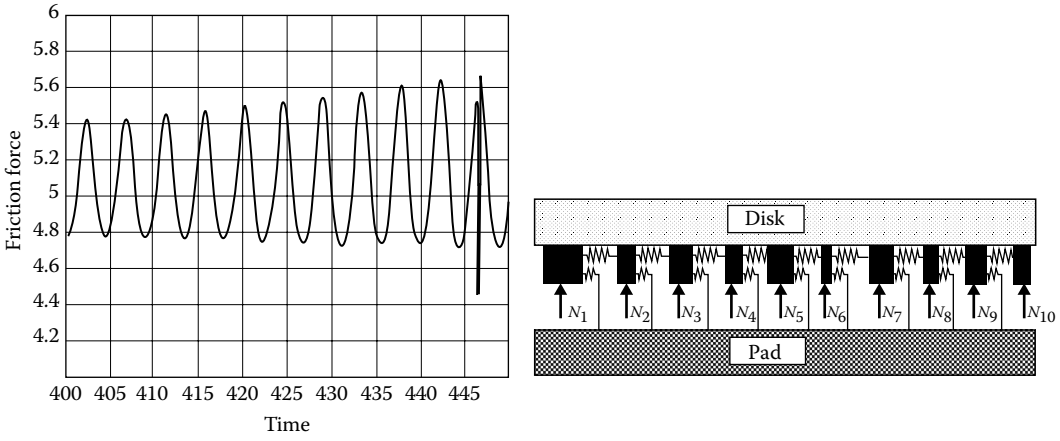


FIGURE 24.35 Variable friction force acting on the disk as a result of many interconnected vibrating masses. (Adapted from Ostermeyer, G.-P., On tangential friction induced vibrations in brake systems, Technical Report, SAE, Paper No. 2008-01-2580, Warrendale, PA, 2008.)

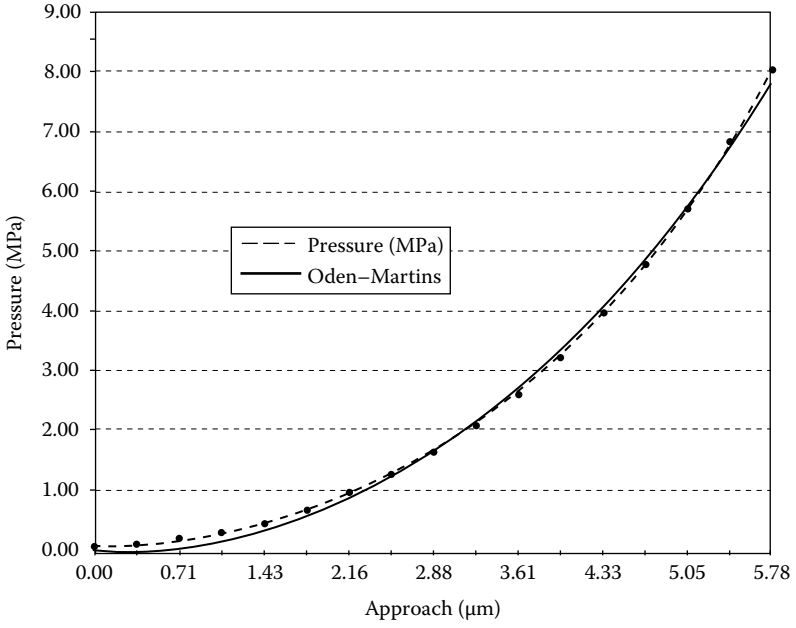


FIGURE 24.36 Increasing of contact pressure as a function of penetration. (Adapted from Chen, F. et al., *Disc Brake Squeal—Mechanism, Analysis, Evaluation, and Reduction/Prevention*, SAE International, Warrendale, PA, December 2005, Chapter 2, ISBN 978-0-7680-1248-4.)

In [40], Chapter 2 deals with surface topography and surface asperity deformation. A correlation between asperity-based contact models and phenomenological contact models is given. The Oden–Martins constitutive model for contact and friction is introduced that is based on normal compliance law. Such a model is very effective for capturing the nonlinear effect of increasing contact pressure with penetration (Figure 24.36).

24.6.4.2 Lumped-Parameter Models

The first attempts in understanding the causes of brake noise have been made through the study of simple analytical few DOF system models, called "minimal models" or "lumped-parameter models"

(Jarvis and Mills [88], Millner [89], Earles and coworkers [90–94], Nishiwaki [57,95], Brooks [96], Hultén and Flint [97], El-Butch and Ibrahim [32,98], Rudolph and Popp [99,100,], Popp et al. [101], Chowdhary [102,103], Hoffmann and Gaul [104]).

Such models are approximations of the actual system and aggregate components seen as discrete elements, like rigid bodies connected by springs and dampers. Literature is full of examples of minimal models ([25,36]) in particular for the study of brake squeal noise. A key issue concerns the information that we can extrapolate from these models and reliability of the results.

Basically, in [104], two possible categories of models were suggested: (1) analysis models for the understanding of vibration phenomena and (2) models derived from large-scale systems. The first group includes models that are used to reveal the physical mechanisms that lead to the brake vibration and/or noise, such models usually can give only qualitative and phenomenological results. Instead, the latter group includes models that have the aim to obtain quantitative results (often for control application [105,106]) but the derivation of such models is still an open question. The early academic studies (see Section 24.6.1) can be clearly placed in the first category because they tried to understand and to explain the phenomenon of vibration and/or noise; some models were validated with experimental analysis.

A number of "minimal models" are presented that appear in the literature. For the sake of simplicity, we will only describe, for each model, the general characteristics and, referring to the linearized equations of motion, we will report only the mass matrix [*M*], the damping matrix [*C*], the stiffness matrix [*K*], and related equations of motion:

$$[M]\cdot\underline{\ddot{z}}+[C]\cdot\underline{\dot{z}}+[K]\cdot\underline{z}=\underline{0} \tag{24.65}$$

The most widely studied model (not only within the brake noise problem) is depicted in Figure 24.37, where the single DOF model of friction oscillator is represented.

Many authors (i.e., [84,107,108]) have dealt with this simple model consisting of a mass *m* restrained by a linear spring (stiffness *k*) and by a linear viscous damping element whose constant is *c*. The mass is pressed by a vertical force *F* with a moving belt that runs at a *given* speed *v*. Obviously, we have a single equation of motion:

$$m_{TOT}\ddot{x}+c_{TOT}\dot{x}+k_{TOT}x=0 \tag{24.66}$$

Depending on the physical parameters of the system, this model can be unstable. Matsui et al. [109] proposed a simple model, represented in Figure 24.38, trying to explain the onset of squeal noise based on a friction law with μ depending on *v*.
Matrices of linearized equations of motion are

$$\underline{z}=x \qquad [M]=m \qquad [K]=k \qquad [C]=c-Nb$$

It is clear that the sign of the term $c-Nb$ determines if the system can be asymptotically stable or unstable. For Matsui and Murakami, the basic parameters that are affecting the stability of the system are the damping coefficient *c*, the vertical force *N*, and the friction slope *b*.

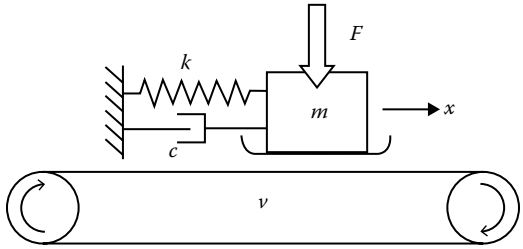


FIGURE 24.37 Single DOF model of friction oscillator.

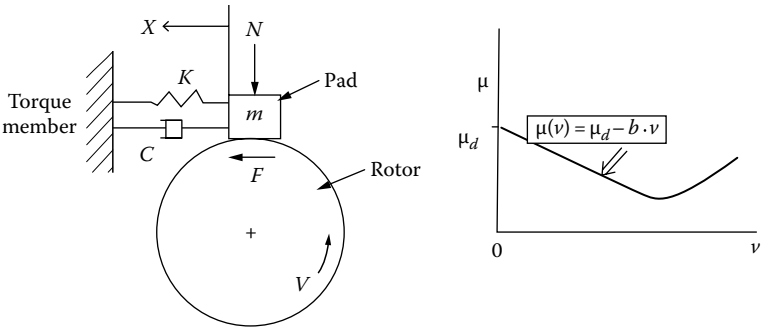


FIGURE 24.38 Model by Matsui and Murakami. (Adapted from Matsui, H. et al., Analysis of disc brake squeal, Technical Report, SAE, Paper No. 920553, Warrendale, PA, 1992.)

Hetzler et al. [110,111] found that the transition behavior between stable and unstable states of sliding motion is more complex than suggested by the classical stability analysis and involves bifurcation (sub- or supercritical) that depends on the system's parameters.

Despite its simplicity and the importance of certain phenomena that can be explained by means of it, the models in Figures 24.37 and 24.38 are definitely not suitable to describe brake vibration and/or noise.

One of the most important "minimal models" in the history of the study of brake squeal is the binary flutter model produced by North in 1972 [112]; such a model was simplified by the same author in 1976 [28]. The model in [28] is a 2-DOF (y and θ) disk, modeled as a rigid body of mass m, thickness $2h$, and moment of inertia J. The disk is assumed to be sandwiched between two layers of friction material, each of length L, of total and stiffness k_1 and $k_2 = k_1L^2/3$. N_0 is the static preload between the pads and the disk and the friction follower forces F_1 and F_2 acting on the disk are (Figure 24.39):

$$F_1 = \mu(k_1 y + N_0), \quad F_2 = \mu(-k_1 y + N_0) \tag{24.67}$$

As mentioned, this model was innovative because it was the first one to consider brake squeal as a self-excited vibration induced by friction forces with constant friction μ. The linearized equations of motion (referring to the steady equilibrium position $y = 0$, $\theta = 0$) are as follows:

$$\underline{z} = \begin{bmatrix} y \\ \theta \end{bmatrix} \quad [M] = \begin{bmatrix} m & 0 \\ 0 & J \end{bmatrix} \quad [K] = \begin{bmatrix} k_t + k_1 & 2\mu N_0 \\ -2\mu h k_1 & k_r + k_2 \end{bmatrix}$$

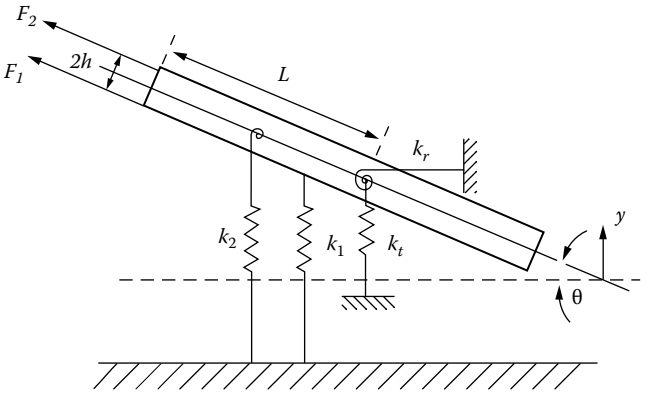


FIGURE 24.39 Model by North. (Adapted from North, M.R., Disc brake squeal, in *Braking of Road Vehicles*, Automobile Division of the Institution of Mechanical Engineers, Mechanical Engineering Publications Limited, London, U.K., pp. 169–176, 1976.)

North observed that the non symmetry of the matrix [*k*] could bring mode-coupling instability or binary flutter. Such an instability occurs if two conditions (Equation 24.68) are verified: The first condition always holds; the second depends on the system's parameter values:

$$(-2\mu h k_1)(2\mu N_0) < 0 \tag{24.68a}$$

$$\frac{1}{mJ}((k_t + k_1)J - (k_r + k_2)m)^2 \le 16\mu^2 h k_1 N_0 \tag{24.68b}$$

Despite the simplicity of this model, which does not include damping, an important result is achieved: Increasing the coefficient of friction μ causes a greater propensity to flutter instability and then to brake vibration and/or noise.

Another important model is the POD model presented by Earles and coworkers: in this model (Figure 24.40), the disk is represented by a rigid body mass *M*; the pad has a rigid body mass *m* and inertia moment J. The two bodies are in contact through a pin placed at characteristics lengths l_C and l_G. Before considering the contact constraint, this model has 3 DOFs (x_d, x_p, φ), but as the pin must remain always in contact with the disk, the model has only 2 DOFs. The coefficient of friction μ in this model is constant.

The equations of motion are

$$\begin{aligned} M\dot{x}_d + Kx_d &= -F \\ m\ddot{x}_p + k_t x_p &= F \\ J\ddot{\varphi} + k_\varphi \varphi &= \mu F l_G + F l_C \end{aligned} \tag{24.69}$$

The contact constraint equation is

$$x_d = x_p + l_C\varphi$$

$$\underline{z} = \begin{bmatrix} x_p \\ \varphi \end{bmatrix} \quad [M] = \begin{bmatrix} M+m & Ml_C \\ \mu l_G M + l_C M & J + \mu l_G l_C M + l_C^2 M \end{bmatrix} \tag{24.70}$$

$$[\boldsymbol{K}] = \begin{bmatrix} K + k_t & Kl_C \\ \mu l_G K + l_C K & k_\varphi + \mu l_G l_C K + l_C^2 K \end{bmatrix}$$

This model is based on the sprag-slip theory (introduced without a consistent mathematical formulation referring to equations of motion by Spurr [113]), and its greatest innovation has been to consider for the first time the rotational motion of the pad and not just its translational motion.

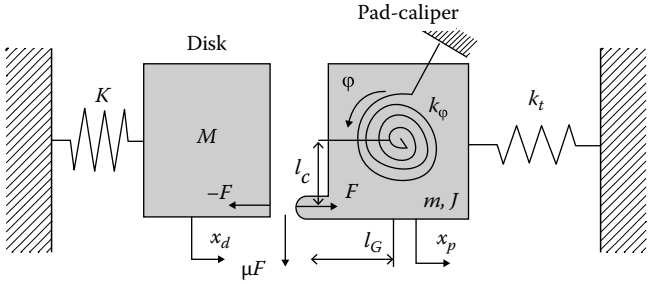


FIGURE 24.40 Model by Earles. (Adapted from Earles, S.W.E., A mechanism of disc-brake squeal, Technical Report, SAE, Paper No. 770181, Warrendale, PA, 1977.)

With a series of models, Earles et al. [90–94] were able to show the importance that the pad bending vibration has on the onset of brake vibration and/or noise. The analysis showed that squealing is generated when the natural frequency of the disk (f_d) and the pad-caliper (f_φ) are close to each other. However, the experimental results were not in good agreement with the results because, among other reasons, the pin–disk model cannot simulate the restoring moment in the rotational direction at the contact surface but only the reaction force in the translational direction. For this reason, Oura et al. [114], in a very recent work, developed a surface-contact model where the influence of the frictional contact surface (modeled as a distributed spring) on squealing has been examined. In this way, a better agreement with the experimental results has been obtained.

An important minimal model is Millner's [89], Figure 24.41. It is composed of a disk, a single brake pad, and half of the caliper. The model has 6 DOFs that are reduced to five including a constraint equation. It is characterized by a constant friction coefficient μ and based on the sprag-slip theory. This model is similar to North's one. With Millner's model, there is a greater instability likelihood increasing friction coefficient μ, but a minimum value for the instability onset is necessary. Other important results from [89] can be summarized as follows. For a constant friction value, the occurrence of squeal and squeal frequency depends on the stiffness of pad material (Young's modulus), caliper stiffness, and mass. Moving the center of pressure, between the piston and the brake pad, toward the leading edge of the brake pad reduced the propensity for squeal. A comparison between the minimal model and the experimental results has been conducted. The theoretical and the real μ values corresponding to squeal appearance are favorably comparable and an increase, at higher values, of the squeal amplitude with μ can be also noted in the model [89].

Murakami et al. developed a more complex model of a pin slider fist-type disk brake [115]. The torque members (at the trailing and leading side of the brake) act as a stiffness/damper system. In relation to Figure 24.42, the 7 DOFs of the model are represented by the rotations θ_o, θ_i,

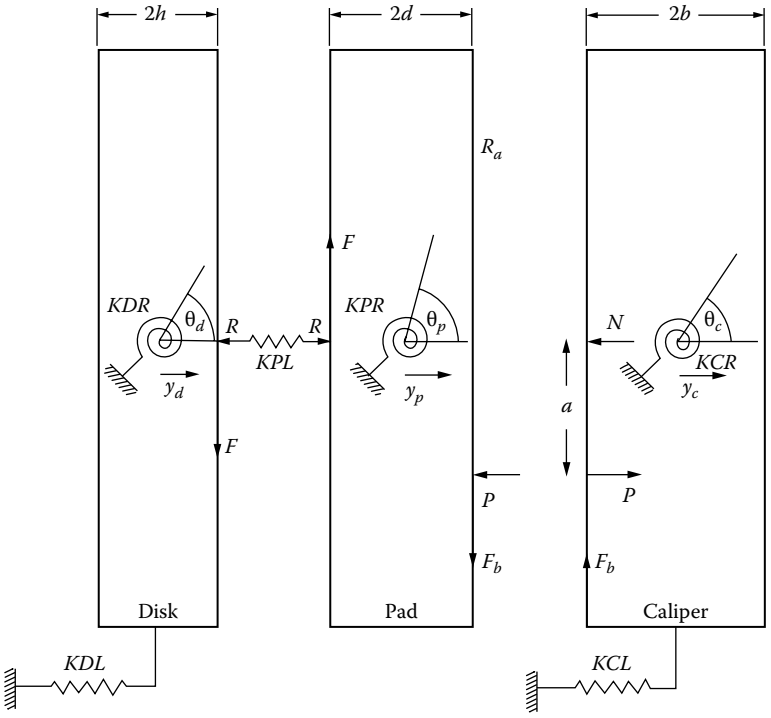


FIGURE 24.41 Model by Millner. (Adapted from Millner, N., An analysis of disc brake squeal, Technical Report, SAE, Paper No. 780332, Warrendale, PA, 1978.)

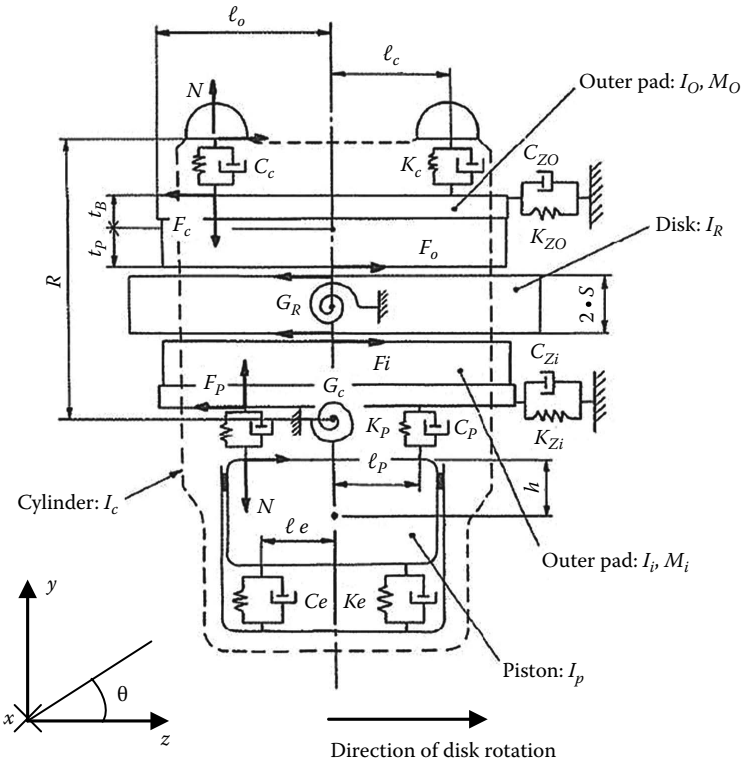


FIGURE 24.42 Model by Murakami, Tsunada, and Kitamura. (Adapted from Murakami, H. et al., A study concerned with a mechanism of disc-brake squeal, Technical Report, SAE, Paper No. 841233, Warrendale, PA, 1984.)

θ_R, θ_p, and θ_c and the displacements in the rotational directions Z_o and Z_i. The authors tried to explain the brake squeal phenomenon both through experimental and numerical methods (FEM and analytical). The main squeal energy source (in relation to the experimental results) is to be identified in the μ–v characteristic of the pad (Figure 24.43a) and in the braking force (acting

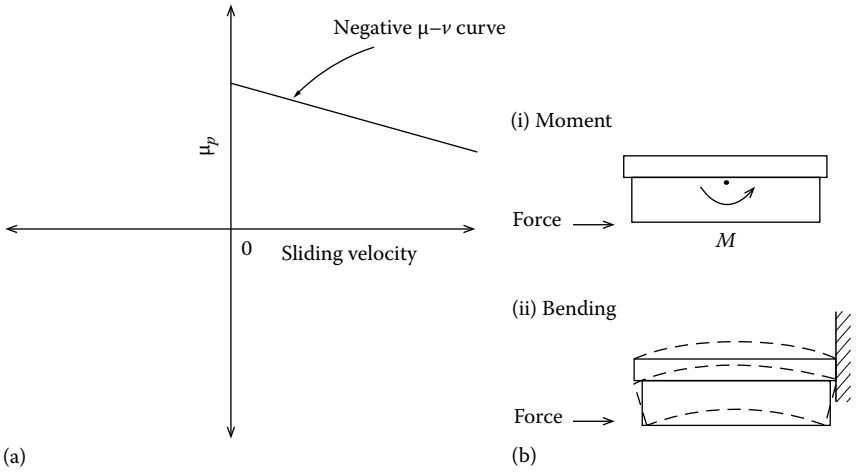


FIGURE 24.43 μ–v characteristic assumed in the model of Figure 24.15. (Adapted from Murakami, H. et al., A study concerned with a mechanism of disc-brake squeal, Technical Report, SAE, Paper No. 841233, Warrendale, PA, 1984.)

along the tangential direction to the disk rotation, see Figure 24.43b, (i) and (ii)) that causes a bending moment in the plane normal to the disk rotation. To confirm the reliability of the calculations, brake tests have been conducted in order to verify the lumped model experimentally and to understand the influence of each parameter on the rate of squeal occurrence. A good agreement was claimed. Given the following parameters:

- *TC*, *TP*, *TE*, *TCR*, *TPR*, *GRO*, and *GRI* (the moments due to the stiffness forces)
- *DC*, *DP*, *DE*, *DCR*, *DPR*, *DRO*, and *DRI* (the moments due to the damping forces)
- *KZO*, *KZI* (*CZO*, *CZI*) (the stiffness (damping) forces)
- *N* (the force along the axis *y*)
- F_o, F_i, F_c, and F_p (the friction forces)

the equations of motion are

$$\begin{aligned}
I_o\ddot{\theta}_o &= -TC - DC + F_o \cdot t_p + F_c \cdot t_B + N \cdot l_c - GRO - DRO \\
I_R\ddot{\theta}_R &= -G_R\theta_R - C_R\dot{\theta}_R + F_o \cdot S - F_i \cdot S - TCR - TPR - DCR - DPR \\
I_i\ddot{\theta}_i &= -TP - DP - F_i \cdot t_p - F_p \cdot t_B - N \cdot l_p - GRI - DRI \\
I_p\ddot{\theta}_p &= TP + DP - TE - DE + N \cdot l_p - F_p \cdot h \\
I_c\ddot{\theta}_c &= -G_c\theta_c + TC + DC + TE + DE - N \cdot l_c - F_c \cdot R \\
M_o\ddot{Z}_o &= -KZO - CZO + F_o - F_c \\
M_i\ddot{Z}_i &= -KZI - CZI + F_i - F_p
\end{aligned} \tag{24.71}$$

Another interesting mathematical model was developed by Nishiwaki to compare the equations of motion for a drum brake squeal, disk brake squeal, and disk brake groan noise [95]. Such brake noises are all generated by the dynamic instability in friction forces that act as nonconservative forces. Vibrations on the brake structures cause the displacements of the friction surfaces and, consequently, the friction forces work on the deformed surfaces of the brake system. The theoretical model for the disk brake squeal is shown in Figure 24.44. As in Murakami's model, the brake used in this study is a pin slider fist-type caliper. The disk is a circular plate with radius r_2, plate thickness t_d, and flexural rigidity D_d. The outer and the inner pads are installed at opposite locations and r_1 and r_2 represent, respectively, the radius of curvature of the inner and outer circumference of each brake pad (t_p is the thickness of the brake pads, D_p their flexural rigidity, and 2ϕ the center angle). The friction material is represented with a lining of thickness h adhered to each brake pad and modeled with a spring without mass (the spring coefficient E_1/h per unit area is distributed over the all friction surface).

In relation to the model's scheme (Figure 24.44), 4 DOFs are possible. Both the deflection of the disk within its plane, v_d, and the deflection of the brake disk and pads along the plane normal to previous one, w_{op}, w_{ip}, w_d, are modeled. The Lagrange equations of motion (expressed in a general form) are

$$M\ddot{X} + KX = \left(\frac{E_1}{h}\right)AX + \mu\left(\frac{E_1}{h}\right)BX \tag{24.72}$$

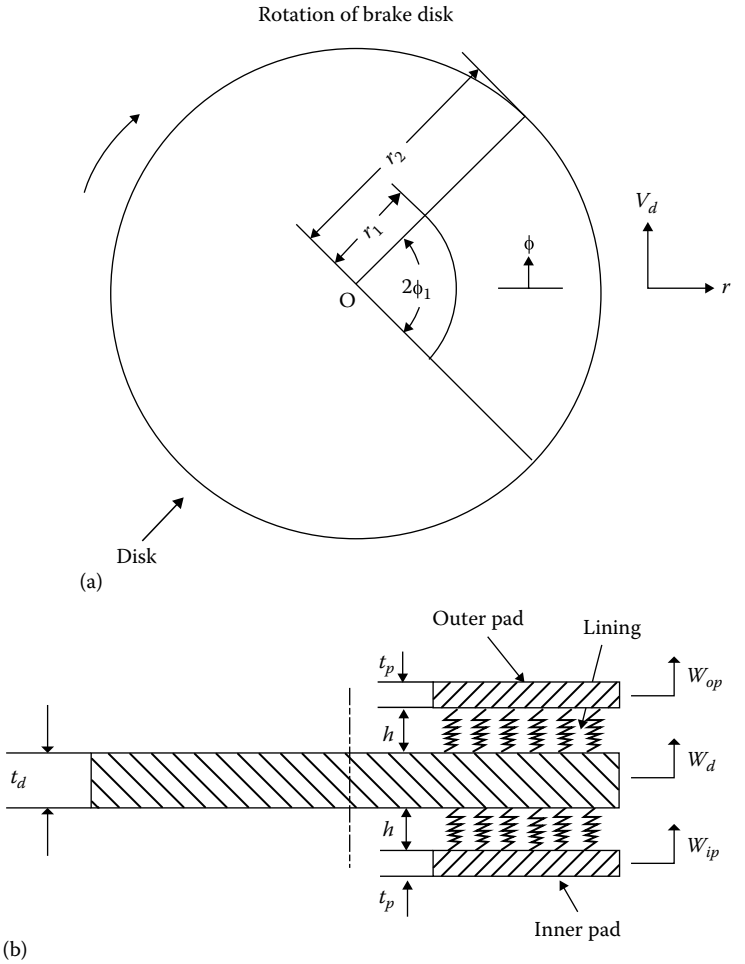


FIGURE 24.44 Model by Nishiwaki. (Adapted from Nishiwaki, M., *Proc. Inst. Mech. Eng.*, 207, 195, 1993.) (a) Plane view. (b) Section.

where, defining a_{ij} and b_{ij} as the constants depending on the vibration modes of the brake pad and disk, the brake pad shape, the lining position, and the brake pad, the terms M, K, A, B, and X assume the following form

$$X = \begin{bmatrix} q_{op}(t) \\ q_{ip}(t) \\ q_{d1}(t) \\ q_{d2}(t) \end{bmatrix}; \quad M = \begin{bmatrix} m_{op} & 0 & 0 & 0 \\ 0 & m_{ip} & 0 & 0 \\ 0 & 0 & m_{d1} & 0 \\ 0 & 0 & 0 & m_{d2} \end{bmatrix}; \quad K = \begin{bmatrix} k_{op} & 0 & 0 & 0 \\ 0 & k_{ip} & 0 & 0 \\ 0 & 0 & k_{d1} & 0 \\ 0 & 0 & 0 & k_{d2} \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} \\ & a_{22} & a_{23} & a_{24} \\ & & a_{33} & a_{34} \\ & & & a_{44} \end{bmatrix}; \quad B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ b_{31} & b_{32} & 0 & 0 \\ b_{41} & b_{42} & 0 & 0 \end{bmatrix} \tag{24.73}$$

Nishiwaki's models are limited to a pure mathematical study of the brake noise phenomena so that an experimental validation has not been conducted.

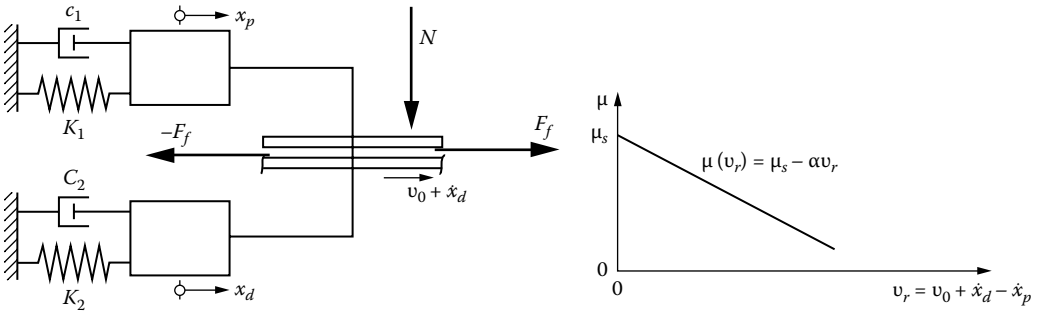


FIGURE 24.45 Model by Shin. (Adapted from Shin, K. et al., *JSME Int. J. C*, 45(2), 426, 2002.)

Brooks et al. developed a lumped-parameter model of brake squeal (with 12 DOFs). Such as in Murakami's model, Brooks considered the vibration in a disk brake as a friction-induced instability caused by a friction force that is a follower force. In addition, the authors used an eigenvalue sensitivity analysis to analyze the effects of small variations in the model parameters on the occurrence of squeal.

Shin's model [116] try to explain self-excited vibrations by a falling friction characteristic and stick-slip theory. The model has the advantage of being very simple, as well as the friction law that is linear, but it is difficult to link it to a real brake system. In fact, an experimental validation of the analytical results has not been made. μ*(v)* law and not linear motion equations are depicted in the succeeding text (Figure 24.45).

$$\begin{aligned} m_1\ddot{x}_p + c_1\dot{x}_p - N\alpha(\dot{x}_p - \dot{x}_d) + k_1 x_p &= N(\mu_S - \alpha v_0) \\ m_2\ddot{x}_d + c_1\dot{x}_d - N\alpha(\dot{x}_d - \dot{x}_p) + k_2 x_d &= -N(\mu_S - \alpha v_0) \end{aligned} \tag{24.74}$$

The linearized equations of motion are

$$\underline{z} = \begin{bmatrix} x_p \\ x_d \end{bmatrix}; \quad [M] = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}; \quad [C] = \begin{bmatrix} c_1 - N\alpha & N\alpha \\ N\alpha & c_2 - N\alpha \end{bmatrix}; \quad [K] = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \tag{24.75a}$$

The addition of damping to either the disk or the pad alone may make the system more noisy.

A very interesting model from a scientific perspective is the one produced by Hoffmann and Gaul [104], which gathers all of the feature presented in previously presented minimal models. Such a model has the disadvantage to be rather different from an actual brake system, but apparently, the authors wanted to explore and explain the mode-coupling instability as a cause of self-excited friction-induced oscillations. The model in Figure 24.46 consists of a conveyor belt and a massive block with 2 DOFs.

Considering small perturbations around the steady sliding state and approximating the friction force by $F_F = \mu k_3 y$, the authors obtained the following linearized equations of motion

$$\underline{z} = \begin{bmatrix} x \\ y \end{bmatrix}; \quad [M] = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}; \quad [K] = \begin{bmatrix} k_{11} & k_{12} - \mu k_3 \\ k_{21} & k_{22} \end{bmatrix} \tag{24.75b}$$

where

$$k_{11} = k_1 \cos^2 \alpha_1 + k_2 \cos^2 \alpha_2$$

$$k_{12} = k_{21} = k_1 \sin \alpha_1 \cos \alpha_1 + k_2 \sin \alpha_2 \cos \alpha_2$$

$$k_{22} = k_1 \sin^2 \alpha_1 + k_2 \sin^2 \alpha_2 + k_3$$

The friction force acts like a cross-coupling force and flutter instability is possible. The model is not correlated to a specific brake system but, as the authors support in [104], it could "be of considerable help for practical design purpose."

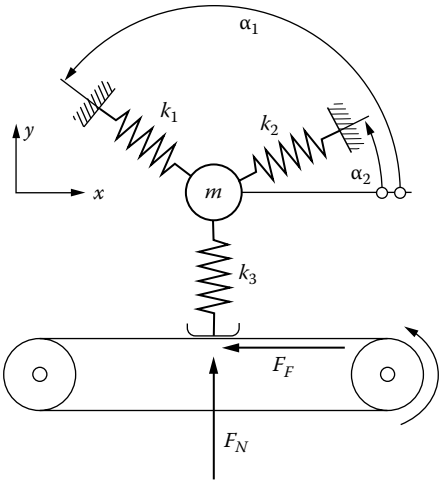


FIGURE 24.46 Model by Hoffmann and Gaul. (Adapted from Hoffmann, N. et al., *Mech. Res. Commun.*, 29, 197, 2002.)

Popp's model [99–101] can be considered a good development of the one by North. Like its inspiring model, it has only 2 DOFs and constant coefficient μ but has been enriched with two springs that simulate the contact pads of the disk and the damping effect on the 2 DOFs of the disk (Figure 24.47).

The linearized equations of motion (see Equation 24.76) may consider a preload N_0 on the two pads to ensure contact:

$$\underline{z} = \begin{bmatrix} x \\ \varphi \end{bmatrix}; \quad [M] = \begin{bmatrix} m & 0 \\ 0 & J \end{bmatrix}; \quad [C] = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$$

$$[K] = \begin{bmatrix} c_1 + c_3 & -c_3 s \\ -c_3(s - \mu h) & c_2 + c_3(s^2 + \mu h s) \end{bmatrix} \tag{24.76}$$

$$[K] = \begin{bmatrix} c_1 + c_3 & -c_3 \\ -c_3(s - \mu h) & c_2 + c_3(s^2 + \mu h s) + 2N_0\left(h(1+\mu^2) + \mu s\right) \end{bmatrix}$$

Flint and Hultén [117] developed a model of a disk brake based on an eigenvalue analysis. In [117], they wrote "implement the disk and the pads as flexible components and incorporate the caliper as a rigid body but not weightless member" in order to investigate the tendency of small vibrations to grow.

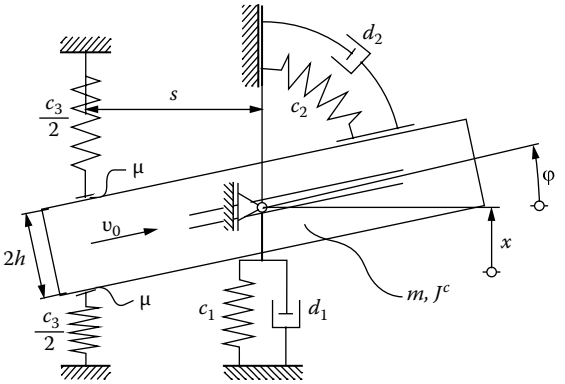


FIGURE 24.47 Model by Popp. (Adapted from Popp, K. et al., *VDI-Bericht*, 1736, 1, 2002.)

The disk is modeled as a flexible beam with cyclo-symmetric boundary conditions and is connected to two pads by the friction material (modeled as "an infinite number of springs"). The model is able to consider the moment created by the unequal friction forces on each side of the disk due to the different compression of the two friction materials. In this way, the rotation of the disk due to the moment can generate a traveling wave as the transverse vibration occurs. The backplates of the pad are represented by beams and they present a flexible contact with ground. The pistons are modeled as rigid bodies (with 1 DOF in the lateral direction), and they are in contact with the backplates by means of seven springs representing in this way a distributed contact (see Figure 24.48).

The equations for a linear and continuous system with external excitation are

$$M[\ddot{w}] + L[w] = 0$$

$$k_{ij} = \int_D \phi_i L[\phi_j] dD \quad m_{ij} = \int_D \phi_i M[\phi_j] dD \tag{24.77}$$

where L and M are linear differential operators; the partial differential equation can be discretized to a system of ordinary equations so that

$$k_{ij} = \int_D \phi_i L[\phi_j] dD \quad m_{ij} = \int_D \phi_i M[\phi_j] dD \tag{24.78}$$

where

D is the domain of integration
ϕ are the test functions ($i, j = 1, 2, \ldots, N$)
N the number of terms included in the analysis (for more details see [117])

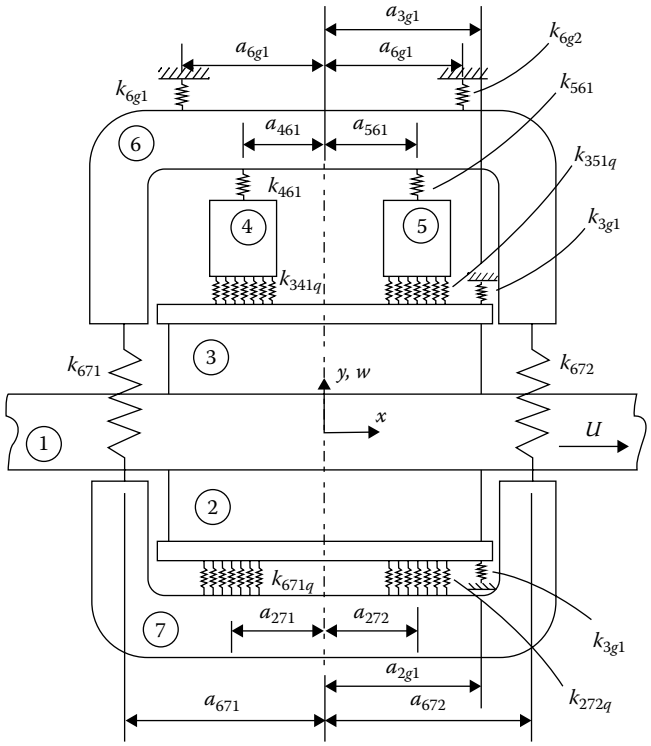


FIGURE 24.48 Model by Flint and Hultén. The model presents a disk (1) and the two pads (2 and 3) as continuous elements, while piston (4 and 5) and the caliper parts (6 and 7) are discrete elements. (From Flint, J. and Hultén, J., *J. Sound Vib.*, 254(1), 1, 2002.)

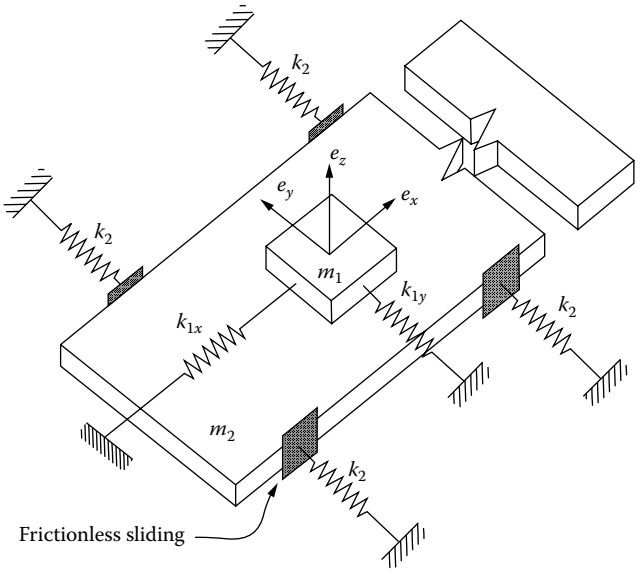


FIGURE 24.49 Model by Kinkaid, O'Reilly, and Papadopoulos. (Adapted from Kinkaid, N.M. et al., *J. Sound Vib.*, 287, 901, 2005.)

The model is able to reproduce the measured squeal frequencies fairly well for a certain real disk brake. The authors supported that "the good agreement between measured squeal and calculated instability is probably due to the fact that these particular squeals are close to disc resonance."

Kinkaid et al. model [24] is a simple model with four translational DOFs that reproduce the sliding of the pad of a disk brake. This model is able to capture the motion of the system in transverse (radial) direction and its main purpose is to simulate the transient phenomena during squealing (Figure 24.49).

In this model, the Amontons–Coulomb frictional law is used and the frictional behavior differs depending on whether the pad is sticking or slipping on the disk. The equations of motion of the system are

$$\begin{aligned} m_1\ddot{x}_1 + c_{1x}\dot{x}_1 + k_{1x}x_1 &= F_{fx} \\ m_1\ddot{y}_1 + c_{1y}\dot{y}_1 + k_{1y}y_1 &= F_{fy} \\ m_2\ddot{x}_2 &= -F_{fx} \\ m_2\ddot{y}_2 + c_{2y}\dot{y}_2 + k_{2y}y_2 &= -F_{fy} \end{aligned} \tag{24.79}$$

where

F_{fx} and F_{fy} represent the x and y components of the friction force
c_{1x}, c_{1y}, and c_{2y} represent the linear viscous damping coefficients

The model conceived by von Wagner et al. [118] has 3 DOFs that are reduced, by means of constraint equations, to two degrees and it reproduces the disk and both of the pads. One of the major advantages of this model is that it can be easily associated with a real disk brake (as one can see in Figure 24.50) but a comparison between the analytical results and a real brake system has not been done.

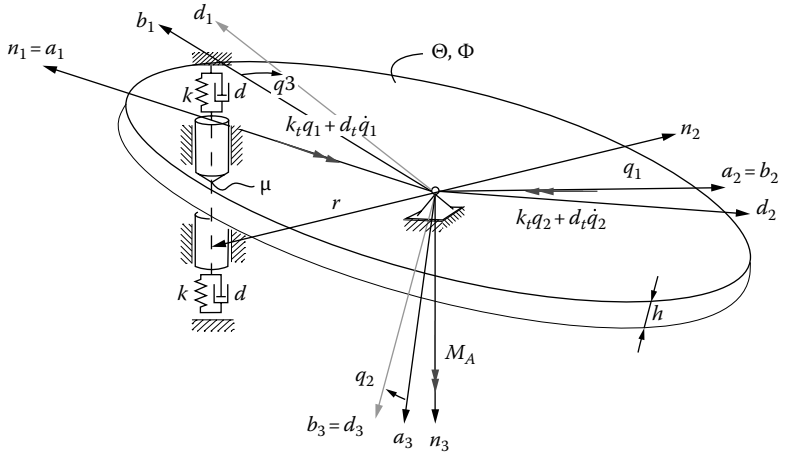


FIGURE 24.50 Model by von Wagner, Hochlenert, and Hagedorn. (Adapted from von Wagner, U. et al., *J. Sound Vib.*, 302, 527, 2007.)

Gyroscopic effects are introduced and the linearized equations of motion are reported in Equation 24.80, μ being the coefficient of friction, r the radius where the pressure is applied, and N_0 the preload:

$$\underline{z} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}; \quad [M] = \begin{bmatrix} \Theta & 0 \\ 0 & \Theta \end{bmatrix}; \quad [C] = \begin{bmatrix} \frac{1}{2}\mu N_0 \frac{h^2}{r\Omega} + 2dr^2 + d_t & \Phi\Omega \\ -\Phi\Omega - \mu dhr & d_t \end{bmatrix}$$
$$[K] = \begin{bmatrix} k_t + 2kr^2 + N_0 h & \frac{1}{2}\mu N_0 \frac{h^2}{r} \\ -\mu(khr + 2N_0 r) & k_t + (1+\mu^2)N_0 h \end{bmatrix} \tag{24.80}$$

North's 8-DOF model [112] is one of the first analytical model well correlated with experimental results and its success was due to the correlation of increased damping and friction coefficient μ to decreasing squeal (see also [36]).

24.6.4.3 Models Derived from a Modal Approach

The main contributions appear to be given by Chan and Mottershead [119–122], Ouyang [72,123], and Giannini and Sestieri [124]. Here, following [36], a summary of the dynamic properties of vibrating disks is briefly reported.

Figure 24.51 represents a vibrating disk, the displacement of a generic point is $\boldsymbol{u} = \boldsymbol{u}(R, \phi, z, t)$; $\boldsymbol{u} = u_x \boldsymbol{e}_r + u_y \boldsymbol{e}_\phi + u_z \boldsymbol{E}_z$. The partial differential equation that governs the motion in the z-direction is

$$\rho h \frac{\partial^2 u_z}{\partial t^2} = \frac{h}{R}\frac{\partial}{\partial R}\left(\sigma_{RR} R \frac{\partial u_z}{\partial R}\right) + \frac{h\sigma_{\phi\phi}}{R^2}\frac{\partial^2 u_z}{\partial \phi^2} - \frac{Eh^3}{12(1-\nu^2)}\nabla^4 \boldsymbol{u}_z + q \tag{24.81}$$

where

h is the thickness of the brake rotor R with ϕ and z are a cylindrical polar coordinate system
u_z is the displacement in z-direction
E is Young's modulus
ρ is the mass density
ν is the Poisson ratio
σ_{RR} and $\sigma_{\phi\phi}$ are the stresses induced in the circular plate by the rotation Ω

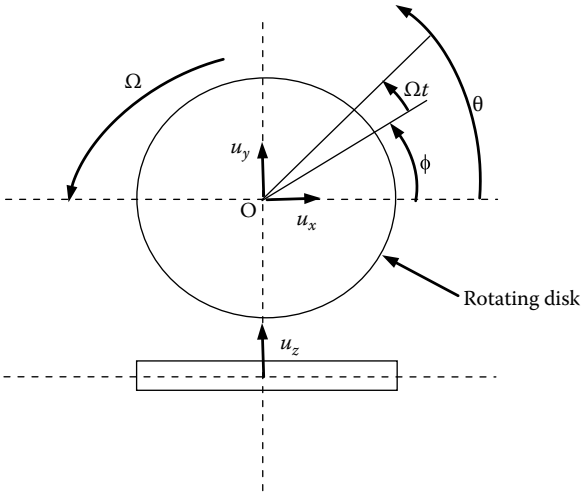


FIGURE 24.51 Reference system for a vibrating disk.

$q = q(R, \phi, t)$ represents the contribution of the external forces and moments and can be expressed as

$$q = -\left(\frac{\delta(R-R^*)\delta(\phi-\phi^*)}{R^*}\right)\frac{F_\phi}{R^*}\frac{\partial u_z}{\partial \phi} \tag{24.82}$$

where

δ is the Dirac delta function

R^* and ϕ^* are the positions where the external force is applied on the disk

Futhermore, ∇^2 is equal to

$$\nabla^2 = \frac{\partial^2}{\partial R^2} + \frac{1}{R}\frac{\partial}{\partial R} + \frac{1}{R^2}\frac{\partial^2}{\partial \phi^2} \tag{24.83}$$

If the external loading is absent and the disk does not rotate around its axis,

$$\rho h \frac{\partial^2 u_z}{\partial t^2} + \frac{Eh^3}{12(1-\nu^2)}\nabla^4 u_z = 0 \tag{24.84}$$

The solution of the above partial differential equation is

$$u_z(R, \phi, t) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} F_{mn}(R)\sin(\omega_{mn}t)\left(A_{mn}\sin(n\phi + \psi_n)\right) =$$

$$= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} F_{mn}(R)\sin(\omega_{mn}t)\left(B_{mn}\sin(n\phi) + C_{mn}\cos(n\phi)\right) \tag{24.85}$$

ψ_n depends on where the excitation is given. The mth modes refer to circumferential nodal lines and are called singlet modes. Diametric nodal lines refer to nth modes and are related to *doublet* modes. Such modes are identical, have the same frequency (ω_{mn}), but their spatial phase is $\pi/2$.

The doublet modes can be interpreted as two waves traveling in opposite direction, actually

$$2\sin(\omega_{mn}t)\sin(n\theta+\psi_n)=\sin(n\theta-\omega_{mn}t+\psi_n)+\sin(n\theta+\omega_{mn}t+\psi_n) \tag{24.86}$$

If the disk is rotating but the centrifugal forces can be neglected, then the asymmetry caused by the disk rotation can be argued

$$2\sin(\omega_{mn}t)\sin(n\theta-n\Omega t+\psi_n)=\sin\left(n\theta-(\omega_{mn}+n\Omega)t+\psi_n\right)+$$
$$+\sin\left(n\theta+(\omega_{mn}-n\Omega)t+\psi_n\right) \tag{24.87}$$

The nodal diameters may rotate at either uniform or nonuniform speed [36,125]. In Figure 24.51, the out-of-plane modes are shown.

By inspection of Tables 24.5 and 24.6 it is clear how an ideal disk can vibrate. The actual disk of a car brake is quite different, due to the so-called hat and bolts acting as constraints. Additionally, the excitation comes from the vibration of the pad(s), which are in contact with the disk destroying the axial symmetry of the mechanical model. The vibration amplitudes are of the same order of magnitude of the disk irregularities (disk thickness variation (DTV), see Section 24.6.3), that is, some micrometer. The role of the excitation of DTV seems often underestimated.

Sestieri in 2006 [124] (and his coworkers) developed a modal model able to predict the squeal frequencies that occur during experiments performed on a laboratory brake setup. The proposed model accounts for the out-of-plane dynamics of the disk (see Figure 24.52). The small brake pad was modeled and its eigenfrequency proved to be crucial for squeal. They reported (see Figure 24.53) that when the eigenfrequency of the pad was nearly equal to the one of a doublet mode, the frequency of the doublet split and instability occurred.

In Table 24.6 the in-plane and combined (out-of-plane and in-plane) modes are shown.

In [37], a section is dedicated to the many contributions that appeared in the literature based on the modal approach.

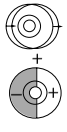
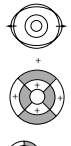
24.6.4.4 FEM Models

Again in [37], a section is dedicated to the many contributions that appeared in the literature based on the FEM approach.

In [126,127], Nack pioneered the derivation of nonlinear FEM analysis of brake vibration and/or noise. FEM modeling was used to describe the nonlinear effects and related transient analysis, but few details were given on how the contact between the pad and the disk was implemented. Actually, the crucial problem is how to connect the element of the disk with the ones of the pad. In [121], apparently, only a point was the location where forces were exchanged between pad and disk. This approach is common to most researches carried out until the beginning of 2000. Additionally, many papers did not account for the rotating disk [37]. Nonetheless, some hints were given on the instability, for example, in [128–130], the role of in-plane vibration of the disk was recognized, thanks to FEM analysis. In the importance of moving, load was pointed out. In [72], the caliper, the carrier, and the pad are modeled by FEM; the disk vibration is modeled analytically. The vibration modes when the disk is still are computed and agree well with experiments. There is a reasonable agreement between computed and measured results, even if, as usual for brake vibration and/or noise investigation, the predicted and experienced squeal frequencies do not always coincide.

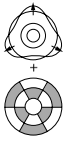
In [131], a full FEM model of a disk brake was presented. At first (Figure 24.54a), a simple pad pressed against a hard body was studied. By exploiting the LS-DYNA code and letting the hard body slide under the contact, setting a constant friction coefficient ($\mu=0.4$), a limit cycle of the center of the contacting surface is computed (Figure 24.54b). Such an analysis leads directly to the computation of an "apparent" friction coefficient that diminishes as the sliding velocity increases justifying the Stribeck effect (see Section 24.6.4.1).

TABLE 24.5
In-of-Plane and Combined Modes of a Brake Disk

Mode	Type m = Number of Nodal Circles n = Number of Nodal Diameters q = Radial Mode Index l = Tangential Mode Index	Frequency Range (Hz)	Shape
	$l = 0$	1900	
	$(m, n) = (1, 1)$	2000	
	$(m, n, q) = (0, 1, 1)$	2500	
1R Radial mode		3000	
2R Radial mode		3200	
	$(m, n) = (1, 2)$	3200	
	$(m, n, q) = (1, 2, 2)$	3800	
1NC1ND		4200	
1NC2ND		4300	
3R Radial mode		5500	

(*continued*)

TABLE 24.5 (continued)
In-of-Plane and Combined Modes of a Brake Disk

Mode	Type m = Number of Nodal Circles n = Number of Nodal Diameters q = Radial Mode Index l = Tangential Mode Index	Frequency Range (Hz)	Shape
BR Breathing mode in radial direction		6000	
	$(m, n, q) = (1, 3, 3)$	7100	
	$q = 0$	7200	

Sources: Shape of the vibrating brake disc adapted from Abu Bakar, A.R. et al., Modelling and simulation of disc brake contact analysis and squeal, PhD thesis, Department of Engineering, University of Liverpool, Liverpool, U.K., 2005; Abu Bakar, A.R. et al., *J. Teknol*., 45(A), 1, 2006; Abu Bakar, A.R. and Ouyang, H., *Int. J. Veh. Des*., 46(4), 416, 2008; Cho, H.-J. and Cho, C.-D., *Proc. IMechE, Part D J. Automob. Eng*., 222, 895, 2008; Giannini, O. et al., *J. Sound Vib*., 292, 1, 2006; Casini, P. et al., *Nonlinear Dyn*., 46, 259, 2006; Giannini, O. and Sestieri, A., *J. Sound Vib*., 296, 583, 2006; Massi, F. et al., *Mech. Syst. Signal Process*., 21, 2374, 2007; Giannini, O. and Massi, F., *J. Sound Vib*., 310, 394, 2008; Jearsiripongkul, T. and Hagedorn, P., *Proc. Appl. Math. Mech*., 4, 93, 2004; Kung, S.-W. et al., Complex eigenvalue analysis for reducing low frequency brake squeal, Technical Report, SAE, Paper No. 2000-01-0444, Warrendale, PA, 2000; Malosh, J.B., Disc brake noise reduction through metallurgical control of rotor resonances, Technical Report, SAE, Paper No. 982236, Warrendale, PA, 1998; Masoomi, M. et al., *Polym. Compos*., 27, 461, 2006; Matsuzaki, M. and Izumihara, T., Brake noise caused by longitudinal vibration of the disc rotor, Technical Report, SAE, Paper No. 930804, Warrendale, PA, 1993; Ouyang, H. et al., *IMechE J. Automob. Eng*., 217(10), 867, 2003; Ouyang, H. et al., *Acta Mech. Sin*., 20(4), 418, 2004; Cao, Q. et al., *Int. J. Numer. Methods Eng*., 61, 1546, 2004; Joe, Y.-G. et al., *Int. J. Automot. Technol*., 9(2), 161, 2008; Trichés Jr., M.J. et al., *ABCM J. Braz. Soc. Mech. Sci. Eng*., XXVI(3), 340, 2004; Lee, H., Modal acoustic radiation characteristics of a thick annular disk, Dissertation of The Ohio State University, Columbus, OH, 2003.

TABLE 24.6
The Main Positive and Negative Aspects of the Two Squeal Noise Analysis Approach

	Tires (Imbalance or Nonuniformities)	
Squeal Noise Analysis Approach	**Advantages and Detections**	**Disadvantages and Inaccuracies**
Complex eigenvalue analysis	• All unstable eigenvalues are found in one run (low computational time with respect to transient analysis). • All unstable vibration modes of the brake system can be predicted (modes with a real part), but in the actual case, not all diverge versus a squeal tendency.	• Limited accuracy in transient states (the method linearizes the nonlinearities) • Nonstationary features, such as material properties (time-dependent), cannot be included. • The real magnitude of a motion cannot be known.
Transient analysis	• The finite amplitude of a limit-cycle motion is known. • Time-varying properties of the brake system (introduced by moving loads) can be taken account.	• Runs occur to find a limit-cycle motion. • Time-domain solutions require a lot of memory storage. • Design changes are hard to perform on the basis of the big quantity of data.

Source: Adapted from Ouyang, H. et al., *Int. J. Veh. Noise Vib.,* 1(3/4), 207, 2005.

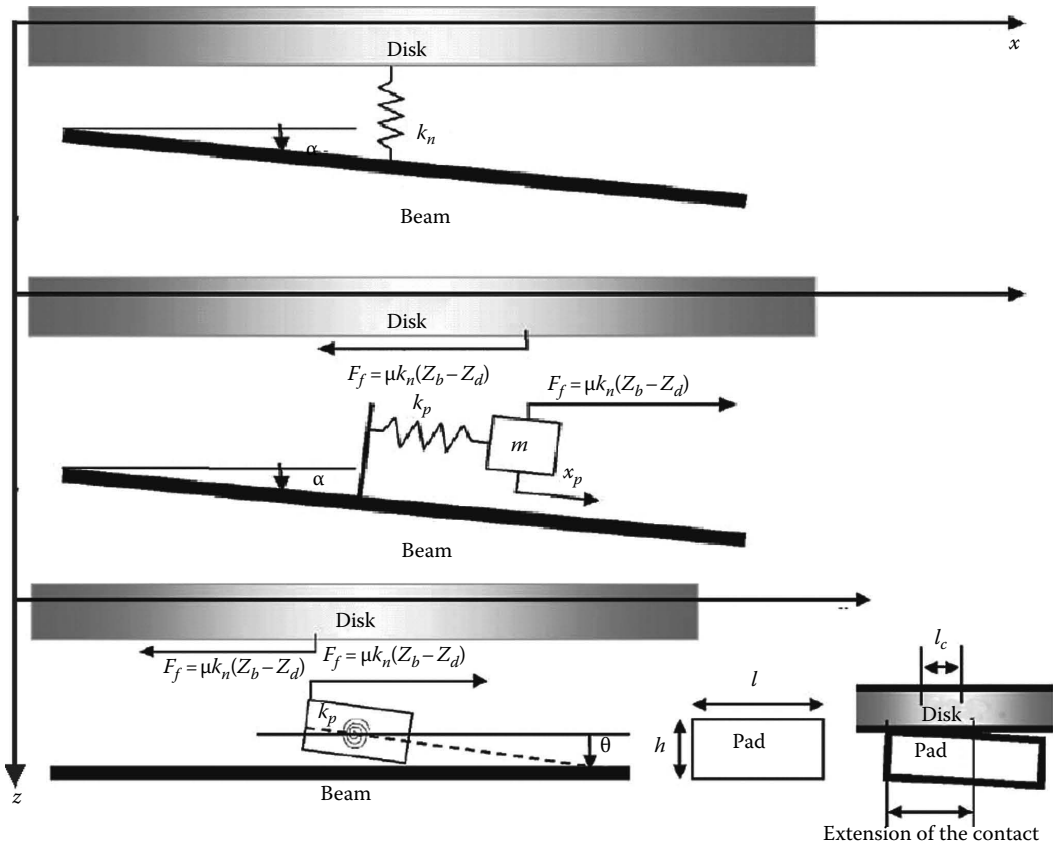


FIGURE 24.52 Out-of-plane, in-plane, and rotational models of the pad. (Adapted from Giannini, O. and Sestieri, A., *J. Sound Vibration*, 296, 583, 2006.)

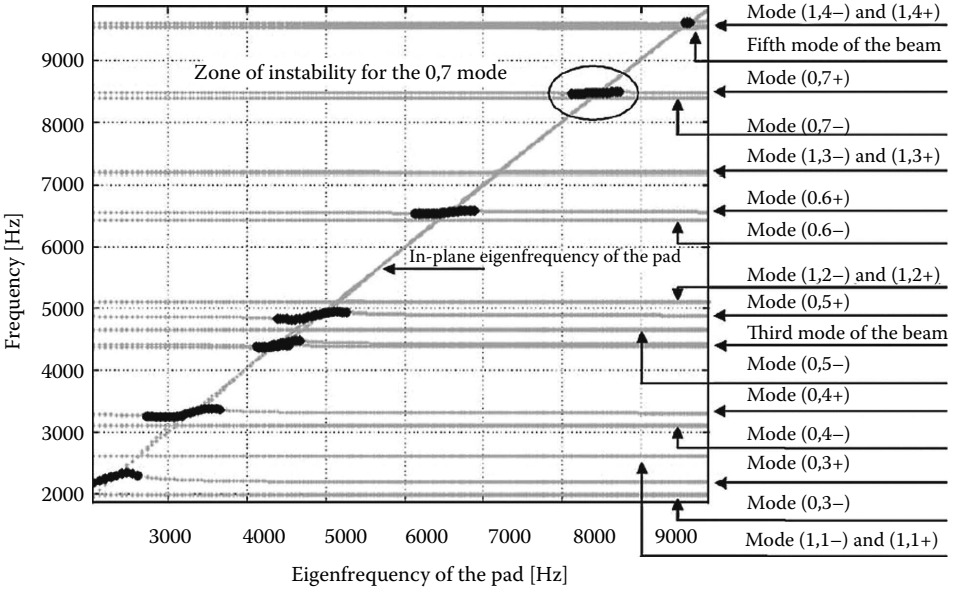


FIGURE 24.53 Example of complex eigenvalue analysis for a given angle of attack of the pad ($\alpha = -1°$). Black lines represent the unstable points where a negative damping generates a squeal occurrence. (Adapted from Giannini, O. and Sestieri, A., *J. Sound Vibration*, 296, 583, 2006.)

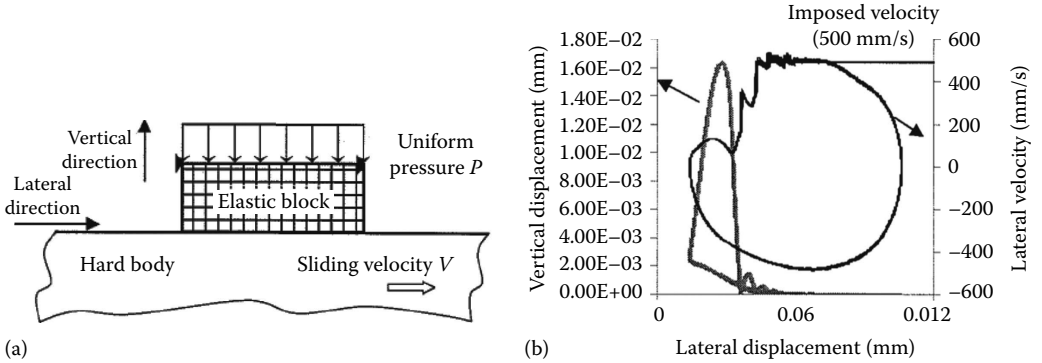


FIGURE 24.54 (a) An elastic block pressed against a sliding hard body and (b) the resulting motion of the center of the contact surface. (Adapted from Brunel, J.F. and Dufrénoy, P., Transient analysis of squealing mode selection in disc brake, Technical Report, SAE, Paper No. 2008-01-2537, Warrendale, PA, 2008.)

24.6.5 Brake Vibration and/or Noise: Causes and Mechanisms

Referring to Figure 24.28, we will focus on the mechanisms that cause brake vibration and/or noise.

24.6.5.1 Judder

Two main groups of judder are identified according to traditional classification [45,46]:

- *Hot judder* caused by thermal deformation (e.g., coning and waving of the disk), uneven thermal expansion, and phase transformation of the disk material
- *Cold judder* caused by geometrical irregularities due to machining and mounting operations, uneven wear, deformation of the bearing system (hub and suspension included), uneven corrosion, or uneven friction film generation

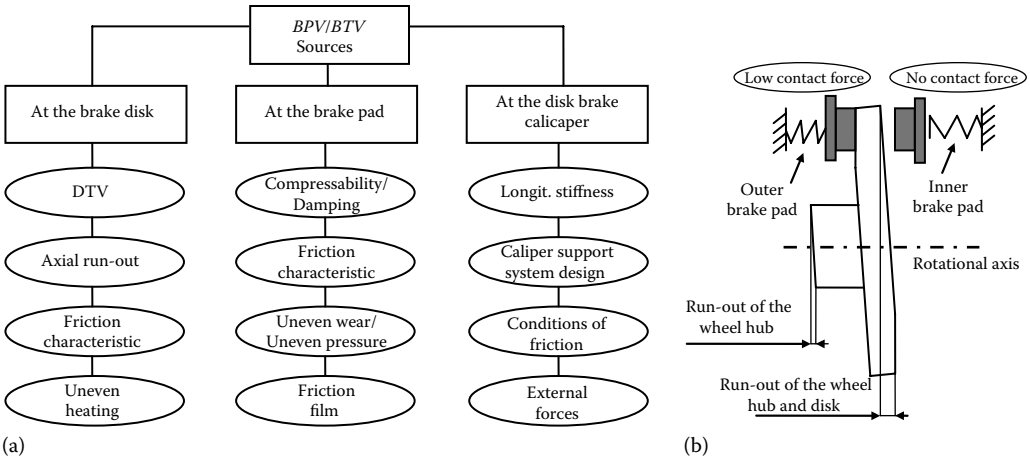


FIGURE 24.55 (a) Causes of brake pressure and brake torque variation (on the left). (Adapted from Jacobsson, H., *Proc. Inst. Mech. Eng., Part D*, 217, 419, 2003.) (b) Erosion on the disk brake surface due to disk run-out during off-brake driving ("cold wear" on the right). (From Grochowicz, J., Experimentelle und theoretische Untersuchungen zu Bremsdruck-und Bremsmomentenschwankungen an Pkw-Scheibenbremsen, Dissertation, Dresden Technical University, Dresden, Germany, 1996.)

Judder can also be classified as a function of its vibration order (many works, such as [45,47], referred to a *low-order geometrical deviations* and a *high-order geometrical deviations*).

The most important factors influencing brake pressure variation (BPV) and brake torque variation (BTV) are summarized in Figure 24.55a [47]. Brake judder is essentially due to geometrical deviations of the brake disk from the ideal shape and spatial variation (along the brake disk) of the coefficient of friction. In particular, the most important geometrical properties are DTV and run-out of the disk (see Figure 24.55b).

The magnitude of the run-out is determined by disk machining and mounting tolerances, bearing clearances and disk distortion during braking. In addition, external forces, such as tire forces and unbalances, affect the run-out. It should be noted that a DTV of 15 μm (measured in cold conditions) is observable as brake judder under hard braking conditions. Therefore, the initial DTV level varies between 6 and 10 μm for most manufacturers in order to guarantee a certain safety margin [132].

Another important effect of judder causing BTV/BPV is the uneven wear of the disk brake. It is caused both during off-brake conditions by geometrical irregularities of the disk brake ("cold erosion") and during braking conditions due to uneven thermal effects [133] (such as uneven contact area and pressure that generate uneven heating distribution on the sliding surface [134,135], uneven friction film thickness that contributes to several micrometer to the thermal growth of the DTV, and uneven corrosion and deposition of heated pad material on the disk).

BTP/BPV generates brake vibrations that are transmitted and amplified to different vehicle components (see Table 24.2). The wheel suspension amplifies the vibrations, especially the frequencies near its eigenfrequencies (for more details, see [136]). The second amplification of the vibration is associated to the steering system. The main ways to reduce judder is to reduce BTV and BPV acting on

- *Pad*: Depending on its tribological properties (density, melting point, flexural and shear strength, compressive stiffness, coefficient of friction and thermal expansion, thermal conductivity, corrosion, and porosity), BTV level can vary by a factor of 2–3 due to the pad composition (with the same DTV). In particular, the pad compressive stiffness should be as low as possible to promote uniform contact pressure and to avoid a thermoelastic instability (hot spots, thermal DTV, etc.), but, at the same time, to achieve a "good pedal feeling,"

the stiffness should be reasonably high and progressive [137,138]. In other words, linear pads are likely to accept higher disk thickness tolerances than 10 μm DTV (brake-by-wire technique offers other ways to achieving a "good pedal feeling") [46,139].

- *Disk*: In addition to the quality of the material (whose main properties are the thermal capacitance, heat dissipation, thermal conductivity, and thermal expansion coefficient), the basic design of the disk rotor is important. Ventilated brake disks are widely used for their weight savings and additional convective heat transfer. However, this type of brake may increase judder problems by inducing an uneven temperature filed around the disk [140]. Nevertheless, the use of composite materials that have particular thermal properties (thermal capacity, thermal expansion, and so on) may be useful to reduce judder and promote uniform contact reducing thermal DTV and hot spots [43–45].

24.6.5.2 Groan, Moan, and Howl

Groan and moan (100–500 Hz) are caused by the vibrations due to the dynamic instability of the wheel–brake system; the same happens for the source of the howl but at higher frequencies (500–1000 Hz).

Groan and moan are caused by stick-slip motion between the friction material and the disk surface. The superposition of several quasi-harmonic vibrations or unsteady vibration impulse sequences of different intensities are its main features. Therefore, this phenomenon is characterized by nonlinear excitation and nonlinear transfer mechanisms.

A very recent and significant model, conceived by von Wagner et al. [118], explains self-excited friction-induced vibrations as a phenomenon caused by the mode-coupling instability and based itself on North's instability criterion.

24.6.5.3 Squeal Noise

In Table 24.7, the many mechanisms for squeal noise generation are summarized.

Stick-Slip: A variable friction coefficient with respect to sliding velocity between pads and rotor is the origin for the brake squeal. When the static friction coefficient is higher than the dynamic coefficient ($\mu_s > \mu_d$), a mass sliding on a moving belt as shown in Figure 24.35 may become unstable. During the starting phase, the sliding force increases until it exceeds the static friction force. Consequently, the mass starts to slide. If the speed drops to zero, the sticking takes place. Then, sliding and sticking occur in succession. Mills in 1938 [141] argued that brake squeal was associated with a decrease in friction coefficient with rubbing speed as shown in Figure 24.35. Nowadays, this mechanism is still thought of as the cause of some low-frequency brake vibration problems.

Sprag-Slip: Spurr [113] introduced the term sprag-slip to name his theory of brake squeal. He actually did not write nor solved a set of differential equations to show instability. His reasoning is sharp but limited to qualitative sentences. The system in Figure 24.55 shows a relatively stiff, massless rod, able to rotate around O. An external force L acts at its free end. Spurr, assuming a friction force at the point contact between the plane and the rod ($F_f = T = \mu_d N$), showed that the system equilibrium implies

$$N = \frac{L}{1 - \mu_d \tan(\theta)}; \quad F_f = \frac{\mu_d L}{1 + \mu_d \tan(\theta)} \tag{24.88}$$

If $\theta \to \tan^{-1}(1/\mu_d)$, then $F_f \to \infty$. The flexibility of the rod let it load and unload, sustaining a *sprag-slip limit cycle* (Figure 24.56).

TABLE 24.7
Brake Squeal Mechanisms

Brake Squeal Mechanism	Authors	References	Figures
Stick-slip	Mills (1938), Bowden and Leben (1939), Blok (1940), Lang and Smales (1983), Eriksson and Jacobson (2001), Shin (2002), Brunel and Dufrenoy (2008)	[50,65,116,131, 141–143]	Figures 24.37, 24.38, and 24.45
Sprag-slip	Spurr (1961), Jarvis and Mills (1963), North (1972/1976)	[28,88,112,113]	Figure 24.39
Stick-slip combined with sprag-slip	Murakami et al. (1984), Earles and coworkers (1971–1987), Felske et al. (1978), Millner (1978)	[89–94,115,144]	Figures 24.41 through 24.43
Mode coupling	North (1972/1976), Millner (1978), Nishiwaki (1993), Hultén and Flint (1999)	[28,89,95,97,112]	Figures 24.39, 24.41, 24.44, and 24.48
Bifurcation (Hopf)	Giannini and Sestieri (2006), Hetzler (2006), Hoffmann and coworkers (2002/2008)	[25,104,124,145]	Figure 24.46
Hammering	Rhee et al. (1989)	[146]	
Coupling of horizontal motions	Kinkaid (2005)	[24]	Figure 24.49
Others	Popp and Stelter (1990), Rudolph and Popp (2001/2002), Mottershead and Chan (1992), Von Wagner (2007), Afferrante and Ciavarella (2008)	[99–101,118,119, 147,148]	Figures 24.47 and 24.50

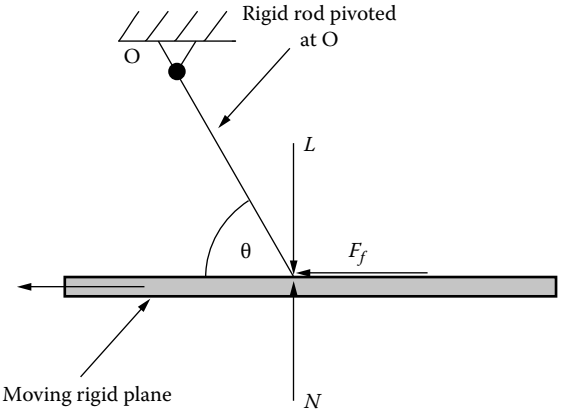


FIGURE 24.56 A stiff, massless rod loaded by an external force L, contacted with a rigid moving plane and pivoted at O. (Adapted from Kinkaid, N.M. et al., *J. Sound Vibration*, 267, 105, 2003.)

Later, researchers included sprag-slip into more complicated (multi-DOFs) models, defining a *geometrically induced* or *kinematic constraint instability*. The variation of the coefficient of friction with sliding speed was not necessary to cause the friction-induced vibrations.

Stick-Slip Combined with Sprag-Slip: Obviously, two mechanisms of instability may be present in the same mechanical model and cause instability. This may happen in models featuring refined and close to reality characteristics.

Mode Coupling: When two modes coalesce (mode locking) instability occurs. Following [78], we report

> The instability mechanism that develops in the laboratory brake can be summarized as follows: the friction force couples the bending vibrations of the laboratory brake and the in-plane vibration of the pad; the friction force is in-phase with the relative displacement between the disc and the pad. If the system vibrates with a frequency close to a natural frequency of the pad, i.e. if a natural frequency of the system falls close to an in-plane eigenfrequency of the pad, the pad vibrates with a 90 deg phase difference with respect to the bending vibration of the disc and the beams. This implies an elliptic trajectory of the contact point on the pad that can transfer energy from the continuous rotation of the disc to the vibrational motion of the pad.

Bifurcation: In nonlinear dynamical systems, the equilibria vary as system parameter varies. When a couple of complex eigenvalues have a positive real part, a Hopf instability occurs. Other saddle-node bifurcation may occur that is still to be assessed as a well-defined squeal mechanism.

Hammering: Rhee et al. proposed [146] a simple mechanical impact model (*hammering*) to explain the excitation mechanism of brake noise. Studies, reported in [146], showed that thermal deformations of the disk brake, as a result of the frictional heat during braking, could cause the formation of localized hot spots on the rotor surface.

Coupling of Horizontal Motions: Such a new mechanism [24] depends crucially on 2-D contact and does not require a friction coefficient that depends on the sliding velocity.

Others: Stick-slip may be correlated to chaos and thus instability. Parametric resonances due to disk vibration can be a source of instability (similar to hammering). The backward waves can be destabilized due to the interaction of the disk and the pad. In [148], thermo elastic dynamic instability (TEDI) is theoretically investigated and foreseen as a cause of brake vibration and/or noise.

24.7 CONCLUSION

In this chapter, the brake system of a vehicle has been described. The most important components of the brake system have been introduced taking into account both their function and the related construction issues.

At first, the functions of a braking system have been introduced, along with some important considerations on basic brake dynamics and energy involved. Then, the brake balance has been considered in-depth.

The hydraulic brake system has been described in detail. The most important construction issues have been briefly addressed. The problem of continuous braking has been considered and the different types of retarders have been described.

In the last section, the fundamental problem of NVH has been deeply investigated by considering all the most important published works. Brake vibration and/or noise has been studied since almost 80 years with minor practical results. This fact suggests that the proper mathematical models have not been identified or employed correctly. Possibly, the available computational power is still reduced to solve the "easy" problem of a pad sliding on a disk. As technology evolved, deeper insight was gained in understanding how pad and disk deform dynamically, which are the actual mechanisms that might be related to vibration and/or noise, which is the role of friction (and what actually friction is), and what are the thermal effects.

For sure, the mechanical/mathematical model of a brake has to be thought of as a complex system, in which many parameters play a crucial role. Such parameters belong either to the meso- or microscale. Some parameters are certain, and some parameters are uncertain. Nonlinearities seems to have a crucial role; actually, common practice suggests the presence of stable and unstable limit cycles.

In a near future, time-domain models will be increasingly used for a priori simulations. The elusive nature of friction will be further investigated. Possibly, the actual irregularities of the disk will have to be taken into account together with the careful mechanical characterization of *all* of the components of a vehicle suspension. The theory of nonlinear systems excited by stochastic processes could prove to be beneficial to assess the probability of occurrence of brake vibration and/or noise.

TABLE 24.8
Appendix: References Sorted by Subject

Brake Noise Topics	Previous	From 1980 to 1990	From 1991 to 2000	Last Years
Reviews	[28,90,149]	[29]	[30,33,41,150]	[35–38,45,151,152]
Low frequency brake noise generation mechanisms			[43,48,126,132, 136,138,153]	[44,46,154]
Squeal noise generation mechanisms	[89,91,92,113,141, 143,155,156]	[57,93,94,115, 146,157]	[31,58,61,95,109, 129,158–164]	[1,24,40,65,165–178,431]
Friction, temperature, and wear	[80,81,91,142,143]	[86,82,146,179]	[59,60,62,63,85,135, 140,180–195]	[25,26,54,64,66–68,71,83,84,100,101,128, 148, 166,196–246]
Vibrational/modal analysis		[147,247,248]	[34,49,96,97,119–122,130, 191,192,194,249–262]	[1,69,73,79,99,107,108,110,111,117,131,145, 154,168,206,208,246, 249,263–315,430]
Complex eigenvalue analysis				
Transient analysis		[429]		
Computational activities	[89,112,141]	[134,316]	[32,98,127,133,139,250, 255,317–324,428]	[24,55,68,72,102–104,114,116,118,123,158, 165,174,198,199,214,215,220,269,271,313, 325–362]
Lumped-parameter, FEM, and control models				
Experimental activities	[144]	[248]	[51,87,125,137, 363,364]	[52,56,78,124,152,365–383]
Both computational and experimental activities			[47,70,74,75,384,385]	[27,53,72,76,77,267,386–403]
Methods to reduce/prevent squeal noise	[155]		[50,181,249,404–409]	[42,101,105,106,228,380, 388,410–425,427]

REFERENCES

1. Breuer, B. and Bill, K.H., *Brake Technology Handbook*, SAE International, Warrendale, PA, 2008, ISBN 978-0-7680-1787-8.
2. Emadi, A., *Handbook of Automotive Power Electronics and Motor Drives*, CRC Press, Boca Raton, FL, 2005, Print ISBN: 978-0-8247-2361-3, eBook ISBN: 978-1-4200-2815-7.
3. Mitschke, M. and Wallwntowitz, H., *Dynamik der Kraftfahrzeuge*, Springer, Berlin, Germany, 2004.
4. Genta, G. and Morello, L., *L'autotelaio. Progetto dei sistemi*, Levrotto & Bella, Turin, Italy, 2007.
5. Morelli, A., *Progetto dell'autoveicolo*, Celid, Turin, Italy, 1999.
6. Limpert, R., *Brake Design and Safety*, 2nd edn., SAE, Warrendale, PA, 1999.
7. Vangi, D., *Ricostruzione della dinamica degli incidenti stradali: Principi e applicazioni*, Firenze University Press, Florence, Italy, 2008.
8. Bosch, *Automotive Handbook*, 7th edn., SAE, Warrendale, PA, 2007.
9. Jazar, R.N., *Vehicle Dynamics: Theory and Application*, Springer Science+Business Media, LLC, New York, 2008, ISBN: 9780387742441.
10. Reimpell, J., Stoll, H., and Betzler, J., *Automotive Chassis: Engineering Principles*, Elsevier Butterworth-Heinemann, Oxford, U.K., 2001.
11. Pacejka, H.B. (eds), *Tire and Vehicle Dynamics*, Butterworth-Heinemann, Oxford, U.K., 2006.
12. Gillespie, T.D., *Fundamentals of Vehicle Dynamics*, SAE, Warrendale, PA, 1992.
13. Genta, G., *Meccanica dell'autoveicolo*, Levrotto & Bella, Turin, Italy, 2000.
14. European Union Directive 71/320/ECC, On the approximation of the laws of the Member States relating to the braking devices of certain categories of motor vehicles and of their trailers, July 26, 1971.
15. Fancher, P. and Winkler, C., Directional performance issues in evaluation and design of articulated heavy vehicles, *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, 45(7–8), 607–647, 2007.
16. Woodrooffe, J., Blower, D., Gordon, T., Green, P.E., Liu, B., and Sweatman, P., Safety benefits of stability control systems for tractor semitrailers, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, 2009.
17. ISO 611, *Road Vehicles—Braking of Automotive Vehicles and Their Trailers—Vocabulary*, 4th edn., April 15, 2003.
18. Heisle, H., *Advanced Vehicle Technology*, Butterworth-Heinemann, Oxford, U.K., 2002.
19. DIN 74000, Hydraulic braking systems; dual circuit brake systems; symbols for brake circuits diagrams, January 8, 1992.
20. Genta, G. and Morello, L., *L'autotelaio. Progetto dei Componenti*, Levrotto & Bella, Turin, Italy, 2007.
21. Garrett, T.K., Newton, K., and Steeds, W., *The Motor Vehicle*, Butterworth-Heinemann, Oxford, U.K., 2001.
22. Orthwein, W.C., *Clutches and Brakes: Design and Selection*, Marcel Dekker Inc., New York, 2004.
23. Pompon, J.P., *Il manuale del disco freno*, Giorgio Nada Editore in collaboration with Brembo, Milan, Italy, 1998.
24. Kinkaid, N.M., O'Reilly, O.M., and Papadopoulos, P., On the transient dynamics of a multi-degree-of-freedom friction oscillator: A new mechanism for disc brake noise, *Journal of Sound and Vibration*, 287, 901–917, 2005.
25. Hoffmann, N.P. and Gaul, L., Friction induced vibration of brakes: Research field and activities, Technical Report, SAE, Paper No. 2008-01-2579, Warrendale, PA, 2008.
26. Ostermeyer, G.-P., On tangential friction induced vibrations in brake systems, Technical Report, SAE, Paper No. 2008-01-2580, Warrendale, PA, 2008.
27. Beloiu, D.M. and Ibrahim, R.A., Influence of random contact forces on disc brake noise using wavelet transform, http://www.aisst.org.tn/PDF/document/sd69.pdf
28. North, M.R., Disc brake squeal, in: *Braking of Road Vehicles*, Automobile Division of the Institution of Mechanical Engineers, Mechanical Engineering Publications Limited, London, U.K., pp. 169–176, 1976.
29. Nishiwaki, M., Review of study on brake squeal, *Japan Society of Automobile Engineering Review*, 11(4), 48–54, 1990.
30. Crolla, D.A. and Lang, A.M., Brake noise and vibration—The state of the art, in: D. Dowson, C.M. Taylor, and M. Godet (Eds.), *Vehicle Tribology*, No. 18 in Tribology Series, Elsevier Science, England, U.K., pp. 165–174, 1991.
31. Ibrahim, R.A., Friction-induced vibration, chatter, squeal, and chaos: Part I: Mechanics of contact and friction, *American Society of Mechanical Engineers Applied Mechanics Reviews*, 47(7), 209–226, 1994.

32. Ibrahim, R.A., Friction-induced vibration, chatter, squeal, and chaos: Part II: Dynamics and modeling, *American Society of Mechanical Engineers Applied Mechanics Reviews*, 47(7), 227–253, 1994.
33. Yang, S. and Gibson, R.F., Brake vibration and noise: Reviews, comments, and proposals, *International Journal of Materials and Product Technology*, 12, 496–513, 1997.
34. Mottershead, J.E., Vibration- and friction-induced instability in disks, *Shock and Vibration Digest*, 30(1), 14–31, 1998.
35. Papinniemi, A., Lai, J.C.S., Zhao, J., and Loader, L., Brake squeal: A literature review, *Applied Acoustics*, 63, 391–400, 2002.
36. Kinkaid, N.M., O'Reilly, O.M., and Papadopoulos, P., Automotive disc brake squeal, *Journal of Sound and Vibration*, 267, 105–166, 2003.
37. Ouyang, H., Nack, W., Yuan, Y., and Chen, F., Numerical analysis of automotive disc brake squeal: A review, *International Journal of Vehicle Noise and Vibration*, 1(3/4), 207–231, 2005.
38. Chen, F., Disc brake squeal: An overview, Technical Report, SAE, Paper No. 2007-01-0587, Warrendale, PA, 2007.
39. Cantoni, C., Cesarini, R., Mastinu, G., Rocca, G., and Sicigliano, R., Brake comfort—A review, *Vehicle System Dynamics, International Journal of Vehicle Mechanics and Mobility*, 47(8), 901–947, 2009.
40. Chen, F., An Tan, C., and Quaglia, R.L., *Disc Brake Squeal—Mechanism, Analysis, Evaluation, and Reduction/Prevention*, SAE International, Warrendale, PA, December 2005, ISBN 978-0-7680-1248-4.
41. Dunlap, K.B., Riehle, M.A., and Longhouse, R.E., An investigative overview of automobile disc brake noise, Technical Report, SAE, Paper No. 1999-01-0142, Warrendale, PA, 1999.
42. Dai, Y. and Lim, T.C., Suppression of brake squeal noise applying finite element brake and pad model enhanced by spectral-based assurance criteria, *Applied Acoustics*, 69, 196–214, 2008.
43. Jacobsson, H., Frequency sweep approach to brake judder, part A: The brake judder phenomenon. Classification and problem approach, Licentiate thesis, Chalmers University of Technology, Göteborg, Sweden, 1998.
44. Jacobsson, H., Brake judder, Thesis, Chalmers University of Technology, Göteborg, Sweden, 2001.
45. Jacobsson, H., Aspects of disc brake judder, *Proceedings of the Institution of Mechanical Engineers, Part D*, 217, 419–430, 2003.
46. Jacobsson, H., Disc brake judder considering instantaneous disc thickness and spatial friction variation, *Proceedings of the Institution of Mechanical Engineers, Part D*, 217, 325–342, 2003.
47. Kao, T.K., Richmond, J.W., and Douarre, A., Brake disc hot spotting and thermal judder: an experimental and finite element study, *International Journal of Vehicle Design*, 23(3/4), 276–296, 2000.
48. Brecht, J., Hoffrichter, W., and Dohle, A., Mechanisms of brake creep groan, Technical Report, SAE, Paper No. 973026, Warrendale, PA, 1997.
49. Bengisu, M.T. and Akay, A., Stability of friction-induced vibrations in multi-degree-of-freedom systems, *Journal of Sound and Vibration*, 171, 557–570, 1994.
50. Lang, A.M. and Smales, H., An approach to the solution of disc brake vibration problems, in: *Braking of Road Vehicles*, Automobile Division of the Institution of Mechanical Engineers, Mechanical Engineering Publications Limited, England, U.K., pp. 223–231, 1993.
51. Ichiba, Y. and Nagasawa, Y., Experimental study on disc brake squeal, Technical Report, SAE, Paper No. 930802, Warrendale, PA, 1993.
52. Fieldhouse, J.D., Steel, W.P., and Talbot, C., The measurement of the absolute displacement of a noisy disc brake, *Proceedings of the IMechE, Part D*, 222, 1121–1140, 2008.
53. Chen, F. and McKillip, D., Measurement and analysis of vibration and deformation using laser metrology for automotive application, *Proceedings of the IMechE, Part D: Journal of Automobile Engineering*, 221, 725–738, 2007.
54. ASTM G 115-04, Standard Guide for Measuring and Reporting Friction Coefficients 2004.
55. Thomsen, J.J. and Fidlin, A., Near-elastic vibro-impact analysis by discontinuous transformations and averaging, *Journal of Sound and Vibration*, 311, 386–407, 2008.
56. SAE J2598, *Brake Pad Damping Measurement Procedure*, U.S. Working Group, 2006. Warrendale, PA 15096-0001, USA.
57. Nishiwaki, M., Harada, H., Okamura, H., and Ikeuchi, T., Study on disc brake squeal, Technical SAE, Paper No. 890864, Warrendale, PA, 1989.
58. Bergman, F., Eriksson, M., and Jacobson, S., Influence of disc topography on generation of brake squeal, *Wear*, 225–229, 621–628, 1999.
59. Bergman, F., Eriksson, M., and Jacobson, S., Surface characterisation of brake pads after running under silent and squealing conditions, *Wear*, 232, 163–167, 1999.
60. Bergman, F., Tribological nature of squealing disc brakes, PhD thesis, Uppsala University, Uppsala, Sweden, 1999.

61. Bergman, F., Eriksson, M., and Jacobson, S., The effect of reduced contact area on occurrence of disc brake squeals for an automotive brake pad, *Proceedings of the Institution of Mechanical Engineers*, *Part D*, 214, 561–568, 2000.
62. Eriksson, M. and Jacobson, S., Tribological surfaces of organic brake pads, *Tribology International*, 33, 817–827, 2000.
63. Eriksson, M., Friction and contact phenomena of disc brakes related to squeal, PhD thesis, Uppsala University, Uppsala, Sweden, 2000.
64. Eriksson, M., Lundqvist, A., and Jacobson, S., A study of the influence of humidity on the friction and squeal generation of automotive brake pads, *Proceedings of the Institution of Mechanical Engineers*, *Part D*, 215, 329–342, 2001.
65. Eriksson, M. and Jacobson, S., Friction behaviour and squeal generation of disc brakes at low speeds, *Proceedings of the Institution of Mechanical Engineers*, *Part D*, 215, 1245–1256, 2001.
66. Eriksson, M., Lord, S., and Jacobson, S., Wear and contact conditions of brake pads: Dynamical in situ studies of pad on glass, *Wear*, 249, 272–278, 2001.
67. Eriksson, M., Bergman, F., and Jacobson, S., On the nature of tribological contact in automotive brakes, *Wear*, 252, 26–36, 2002.
68. Abu Bakar, A.R., Ouyang, H., Titeica, D., and Hamid, M.K.A., Modelling and simulation of disc brake contact analysis and squeal, PhD thesis, Department of Engineering, University of Liverpool, Liverpool, U.K., 2005.
69. Abu Bakar, A.R. and Ouyang, H., A prediction methodology of disk brake squeal using complex eigenvalue analysis, *International Journal of Vehicle Design*, 46(4), 416–435, 2008.
70. Ouyang, H., Mottershead, J.E., Brookfield, D.J., James, S., and Cartmell, M.P., A methodology for the determination of dynamic instabilities in a car disc brake, *International Journal of Vehicle Design*, 23(3/4), 241–262, 2000.
71. Ouyang, H. and Mottershead, J.E., A bounded region of disc-brake vibration instability, *Transactions of the American Society of Mechanical Engineers, Journal of Vibration and Acoustics*, 123(4), 543–545, 2001.
72. Ouyang, H., Cao, Q., Mottershead, J.E., and Treyde, T., Vibration and squeal of a disc brake: Modelling and experimental results, *IMechE Journal of Automobile Engineering*, 217(10), 867–875, 2003.
73. Cao, Q., Ouyang, H., Friswell, M.I., and Mottershead, J.E., Linear eigenvalue analysis of the disc-brake squeal problem, *International Journal for Numerical Methods in Engineering*, 61, 1546–1563, 2004.
74. Schiffner, K. and Rinsdorf, A., Pratical evaluation and sFEM-modelling of a squealing disc brake, Technical Report, SAE, Paper No. 933071, Warrendale, PA, 1993.
75. Tworzydlo, W.W., Hamzeh, O.N., Zaton, W., and Judek, T.J., Friction-induced oscillations of a pin-on-disk slider: Analytical and experimental studies, *Wear*, 236, 9–23, 1999.
76. Beloiu, D.M. and Ibrahim, R.A., Analytical and experimental investigations of disc brake noise using the frequency-time domain, *Structural Control and Health Monitoring*, 13, 277–300, 2006.
77. Rusli, M. and Okuma, M., Squeal noise prediction in dry contact sliding systems by means of experimental spatial matrix identification, *Journal of System Design and Dynamics*, 2(2), 585–595, 2008.
78. Giannini, O., Akay, A., and Massi, F., Experimental analysis of brake squeal noise on a laboratory brake set-up, *Journal of Sound and Vibration*, 292, 1–20, 2006.
79. Massi, F., Baillet, L., Giannini, O., and Sestieri, A., Brake squeal: Linear and nonlinear numerical approaches, *Mechanical Systems and Signal Processing*, 21, 2374–2393, 2007.
80. Kragelskii, V., *Friction and Wear*, Butterworth-Heinemann, Oxford, U.K., 1965.
81. Rabinowicz, E., *Friction and Wear of Materials*, Wiley, New York, 1965.
82. Suh, N.P., *Tribophysics*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
83. Bowden, F.P. and Tabor, D., *The Friction and Lubrication of Solids*, Oxford University Press, Clarendon Press, Oxford, U.K. (second corrected edition, Oxford Classic Text in the Physical Sciences), 2001.
84. Astrom, K.J. and Canudas-de-Wit, C., Revisting the LuGre friction model, *IEEE Control Systems Magazine*, 28(6), 101–114, 2008.
85. Dahl, P.R., Solid friction damping of mechanical vibrations, *AIAA Journal*, 14, 1675–1682, 1996.
86. Ruina, A.L., Constitutive relations for frictional slip, in: Z.P. Bazant (Ed.), *Mechanics of Geomaterials: Rocks, Concrete, Soils*, Wiley, New York, pp. 169–188, 1985.
87. Ibrahim, R.A., Madhavan, S., Qiao, S.L., and Chang, W.K., Experimental investigation of friction-induced noise in disc brake systems, *International Journal of Vehicle Design*, 23(3/4), 218–240, 2000.
88. Jarvis, R.P. and Mills, B., Vibrations induced by dry friction, *Proceedings of the Institution of Mechanical Engineers*, 178(32), 847–857, 1963.
89. Millner, N., An analysis of disc brake squeal, Technical Report, SAE, Paper No. 780332, Warrendale, PA, 1978.

90. Earles, S.W.E. and Soar, G.B., Squeal noise in disc brakes, *Vibration and Noise in Motor Vehicles*, Institution of Mechanical Engineers, London, U.K., Paper No. C 101/71, pp. 61–69, 1971.
91. Earles, S.W.E. and Lee, C., Instabilities arising from the frictional interaction of a pin-disk system resulting in noise generation, *Transactions of the American Society of Mechanical Engineers, Journal of Engineering for Industry*, 98(1), 81–86, 1976.
92. Earles, S.W.E., A mechanism of disc-brake squeal, Technical Report, SAE, Paper No. 770181, Warrendale, PA, 1977.
93. Earles, S.W.E. and Badi, M.N.M., Oscillatory instabilities generated in a double-pin and disc undamped system: A mechanism of disc-brake squeal, *Proceedings of the Institution of Mechanical Engineers*, 198C(4), 43–50, 1984.
94. Earles, S.W.E. and Chambers, P.W., Disc brake squeal noise generation: Predicting its dependency on system parameters including damping, *International Journal of Vehicle Design*, 8(4/5/6), 538–552, 1987.
95. Nishiwaki, M., Generalized theory of brake noise, *Proceedings of the Institution of Mechanical Engineers*, 207, 195–202, 1993.
96. Brooks, P.C., Crolla, D.A., Lang, A.M., and Schafer, D.R., Eigenvalue sensitivity analysis applied to disc brake squeal, in: *Braking of Road Vehicles*, Institution of Mechanical Engineers, Mechanical Engineering Publications Limited, Bury St. Edmunds, England, U.K., pp. 135–143, 1993.
97. Hultén, J.O. and Flint, J., An assumed modes method approach to disc brake squeal analysis, Technical Report, SAE, Paper No. 1999-01-1335, Warrendale, PA, 1999.
98. El-Butch, A.M. and Ibrahim, I.M., Modeling and analysis of geometrically induced vibration in disc brakes considering contact parameters, Technical Report, SAE, Paper No. 1999-01-0143, Warrendale, PA, 1999.
99. Rudolph, M. and Popp, K., Brake squeal: Detection, utilization and avoidence of nonlinear dynamical effects in engineering applications, Final Report of a Joint Research Project Sponsored by the German Federal Ministry of Education and Research, Shaker, Aachen, Germany, pp. 197–225, 2001.
100. Rudolph, M. and Popp, K., Friction induced brake vibrations, *Proceedings of DETC'01* (in CD-ROM), DETC2001/VIB-21509, ASME, Pittsburgh, PA, pp. 1–10, 2001.
101. Popp, K., Rudolph, M, Kröger, M., and Lindner, M., Mechanisms to generate and to avoid friction induced vibrations, *VDI-Bericht*, 1736, 1–15, 2002.
102. Chowdhary, H.V., Modelling of disc brakes systems for squeal prediction, Master's thesis, Purdue University, West Lafayette, IN, 2001.
103. Chowdhary, H.V., Bajaj, A.K., and Krousgill, C.M., An analytical approach to model disc brake system for squeal prediction, *Proceedings of DETC'01* (in CD-ROM), DETC2001/VIB-21560, ASME, Pittsburgh, PA, pp. 1–10, 2001.
104. Hoffmann, N., Fischer, M., Allgaier, R., and Gaul, L., A minimal model for studying properties of the mode-coupling type instability in friction induced oscillations, *Mechanics Research Communications*, 29, 197–205, 2002.
105. Hagedorn, P. and von Wagner, U., "Smart pads": A new tool for the suppression of brake squeal, in: B. Breuer (Ed.), *VDI-Bericht* 575, *Proceedings of XXIV* μ-*Kolloquium*, Bad Neuenahr, Germany, 2003.
106. von Wagner, U., Hochlenert, D., Jearsiripongkul T., and Hagedorn, P., Active control of brake squeal via smart pads, Technical Report, SAE, Paper No 2004-01-2773, Warrendale, PA, 2004.
107. Hetzeler, H., Schwarzer, D., and Seemann, W., Steady-state stability and bifurcations of friction oscillators due to velocity-dependent friction characteristics, *Proceedings of the Institution of Mechanical Engineers*, *Part K*, *Journal of Multibody Dynamics*, 221, 401–412, 2007.
108. Hoffmann, N., Linear stability of steady sliding in point contacts with velocity dependent and LuGre type friction, *Journal of Sound and Vibration*, 301, 1023–1034, 2007.
109. Matsui, H., Murakami, H., Nakanishi, H., and Tsuda, Y., Analysis of disc brake squeal, Technical Report, SAE, Paper No. 920553, Warrendale, PA, 1992.
110. Hetzler, H., Seemann, W., and Seemann, W., Analytical investigation of steady-state stability and Hopf-bifurcations occurring in sliding friction oscillators with application to low-frequency disc brake noise, *Communications in Nonlinear Science and Numerical Simulation*, 12, 83–99, 2007.
111. Hetzler, H., Schwarzer, D., and Seemann, W., Steady-state stability and bifurcations of friction oscillators due to velocity-dependent friction characteristics, *Proceedings of the IMechE*, *Part K*, 221, 401–413, 2007.
112. North, M.R., Disc brake squeal—A theoretical model, Technical Report 1972/5, Motor Industry Research Association, Warwickshire, England, U.K., 1972.
113. Spurr, R.T., A theory of brake squeal, *Proceedings of the Automobile Division of the Institution of Mechanical Engineers*, 1961–1962(1), 33–52, 1961.

114. Oura, Y., Kurita, Y., Matsumura, Y., and Nishizawa, Y., Influence of distributed stiffness in contact surface on disk brake squeal, Technical Report, SAE, Paper No. 2008-01-2584, Warrendale, PA, 2008.
115. Murakami, H., Tsunada, N., and Kitamura, T., A study concerned with a mechanism of disc-brake squeal, Technical Report, SAE, Paper No. 841233, Warrendale, PA, 1984.
116. Shin, K., Oh, J.-E., and Brennan, M.J., Nonlinear analysis of friction induced vibrations of a two-degree-of-freedom model for disc brake squeal noise, *JSME International Journal C*, 45(2), 426–432, 2002.
117. Flint, J. and Hultén, J., Lining-deformation-induced modal coupling as squeal generator in a distributed parameter disc brake model, *Journal of Sound and Vibration*, 254(1), 1–21, 2002.
118. von Wagner, U., Hochlenert, D., and Hagedorn, P., Minimal models for disk brake squeal, *Journal of Sound and Vibration*, 302, 527–539, 2007.
119. Mottershead, J.E. and Chan, S.N., Brake squeal—An analysis of symmetry and flutter instability, in: *Friction-Induced Vibration, Chatter, Squeal, and Chaos*, vol. DE-49, American Society of Mechanical Engineers, New York, pp. 87–97, 1992.
120. Chan, S.N., Mottershead, J.E., and Cartmell, M.P., Parametric resonances at subcritical speeds in discs with rotating frictional loads, *Proceedings of the Institution of Mechanical Engineers*, 208, 417–425, 1994.
121. Mottershead, J.E. and Chan, S.N., Flutter instability of circular discs with frictional follower loads, *ASME Journal of Vibration and Acoustics*, 117, 161–163, 1995.
122. Chan, S.N., Mottershead, J.E., and Cartmell, M.P., Instabilities at subcritical speeds in discs with rotating frictional follower loads, *Transactions of the American Society of Mechanical Engineers, Journal of Vibration and Acoustics*, 117(2), 240–242, 1995.
123. Ouyang, H., Gu, Y., and Yang, H., A dynamic model for a disc excited by vertically misaligned, rotating, frictional sliders, *Acta Mechanica Sinica*, 20(4), 418–425, 2004.
124. Giannini, O. and Sestieri, A., Predictive model of squeal noise occurring on a laboratory brake, *Journal of Sound and Vibration*, 296, 583–601, 2006.
125. Fieldhouse, J.D. and Newcomb, T.P., Double pulsed holography used to investigate noisy brakes, *Optics and Lasers in Engineering*, 6(25), 455–494, 1996.
126. Nack, W.V. and Joshi, A.M., Friction induced vibration: Brake moan, Technical Report, SAE, Paper No. 951095, Warrendale, PA, 1995.
127. Nack, W.V., Brake squeal analysis by finite elements, *International Journal of Vehicle Design*, 23(3/4), 263–275, 2000.
128. Chen, G.X., Zhou, Z.R., Kapsa, P., and Vincent, L., Effect of surface topography on formation of squeal under reciprocating sliding, *Wear*, 253, 411–423, 2002.
129. Matsuzaki, M. and Izumihara, T., Brake noise caused by longitudinal vibration of the disc rotor, Technical Report, SAE, Paper No. 930804, Warrendale, PA, 1993.
130. Bae, J.C. and Wickert, J.A., Free vibration of coupled disk-hat structures, *Journal of Sound and Vibration*, 235(1), 117–132, 2000.
131. Brunel, J.F. and Dufrénoy, P., Transient analysis of squealing mode selection in disc brake, Technical Report, SAE, Paper No. 2008-01-2537, Warrendale, PA, 2008.
132. Engel, H., Bachman, Th., Eichhorn, U., and Saame, Ch., Dynamical behavior of brake-disc geometry as cause of brake judder, *Proceedings of AEEC Fourth International Conference on Vehicle and Traffic System Technology*, vol. 1, pp. 456–481, Strasbourg, France, 1993.
133. Brooks, P.C., Barton, D.C., Crolla, D.A., Lang, A.M., and Schafer, D.R., A study of disc brake judder using a fully coupled thermo-mechanical finite element model, *Proceedings, 25th FISITA Congress*, Beijing, China, pp. 340–349, Technical Report, SAE, Paper No. 945042, Warrendale, PA, 1994.
134. Kao, T.K., Richmond, J.W., and Moore, M.W., The application of predictive techniques to study thermo-elastic instability of brakes, *Proceedings of the Twelfth Annual Brake Colloquium and Engineering Display*, Atlanta, GA, Technical Report, SAE, Paper No. 942087, Warrendale, PA, 1984.
135. Thuresson, D., Thermomechanical analysis of friction brakes, *Proceedings of the Eighteenth Annual Brake Colloquium and Engineering Display*, San Diego, CA, vol. P-358, pp. 149–159, Technical Report, SAE, Paper No. 2000-01-2775, Warrendale, PA, 2000.
136. Gassmann, S. and Engel, H.G., Excitation and transfer mechanism of brake judder, Technical Report, SAE, Paper No. 931880, Warrendale, PA, 1993.
137. Avilés, R., Hennequet, G., Hernández, A., and Llorente, L.I., Low frequency vibrations in disc brakes at high car speed. Part I: Experimental approach, *International Journal of Vehicle Design*, 16(6), 542–555, 1995.
138. de Vries, A. and Wagner, M., The brake judder phenomenon, *SAE Transactions*, 101(6), 652–660, 1992.
139. Richmond, J.W., Kao, T.K., and Moore, M.W., The development of computational analysis techniques for disc brake pad design, *Proceedings of Advances in Automotive Braking Technology* (*Design Analysis and Material Development*), Mechanical Engineering Publication, London, U.K., 1996.

140. Grieve, D.G., Barton, D.A., Crolla, D.A., Chapman, J.L., and Buckingham, J.T., Alternative disc brake materials, *Proceedings of Advances in Automotive Braking Technology: Design Analysis and Material Developments*, Leeds, U.K., 1996.
141. Mills, H.R., Brake squeak, Technical Report 9000 B, Institution of Automobile Engineers, London, U.K., 1938.
142. Bowden, F.P. and Leben, L., The nature of sliding and the analysis of friction, *Royal Society of London Proceedings Series A, Mathematics Physics and Engineering Science*, 169, 371–391, 1939.
143. Blok, H., Fundamental mechanical aspects of boundary lubrication, *SAE Journal*, 46(2), 54–68, 1940.
144. Felske, A., Hoppe, G., and Matthäi, H., Oscillations in squealing disc brakes—Analysis of vibration modes by holographic interferometry, Technical Report, SAE, Paper No. 780333, Warrendale, PA, 1978.
145. Hetzler, H., Seemann, W., and Seemann, W., Bifurcation behavior of a 1DOF sliding friction oscillator, *Proceedings in Applied Mathematics and Mechanics*, 6, 331–332, 2006.
146. Rhee, S.K., Tsang, P.H.S., and Wang, Y.S., Friction-induced noise and vibration of disc brakes, *Wear*, 133, 39–45, 1989.
147. Popp, K. and Stelter, P., Stick-slip vibrations and chaos, *Philosophical Transactions of the Royal Society of London A*, 332, 89–105, 1990.
148. Afferrante, L. and Ciavarella, M.: Thermo-elastic dynamic instability (TEDI)—A review of recent results, *Journal of Engineering Mathematics*, 61, 285–300, 2008.
149. Fosberry, R.A.C. and Holubecki, Z., Interim report on disc brake squeal, Technical Report 1959/4, Motor Industry Research Association, Warwickshire, U.K., 1959.
150. Langthjem, M.A. and Sugiyama, Y., Dynamic stability of columns subjected to follower loads: A survey, *Journal of Sound and Vibration*, 238(5), 809–851, 2000.
151. Chan, D. and Stachowiak, G.W., Review of automotive brake friction materials, *Proceedings of the Institution of Mechanical Engineers Part D*, 218, 953–966, 2004.
152. Chen, F., Luo, W.D., Dale, M., Petniunas, A., Harwood, P., and Brown, G.M., High-speed ESPI and related techniques: overview and its application in the auto motive industry, *Optics and Lasers in Engineering*, 40, 459–485, 2003.
153. Kido, I., Kurahachi, M., and Asai, M., A study on low-frequency brake squeal noise, Technical Report, SAE, Paper No. 960993, Warrendale, PA, 1996.
154. Hetzler, H. and Seemann, W., On low frequency disc-brake vibrations, *Proceedings in Applied Mathematics and Mechanics*, 5, 95–96, 2005.
155. Fosberry, R.A.C. and Holubecki, Z., Disc brake squeal: Its mechanism and suppression, Technical Report 1961/1, Motor Industry Research Association, Warwickshire, U.K., 1961.
156. Harding, P.R.J. and Wintle, B.J., Flexural effects in disc brake pads, *Proceedings of the Institution of Mechanical Engineers*, 192, 1–7, 1978.
157. Fieldhouse, J.D. and Rennison, M., Proposals for the classification of brake noise—An equivalent noise dose rating, Technical Report, SAE, Paper No. 973027, Warrendale, PA, 1997.
158. Chargin, M.L., Dunne, L.W., and Herting, D.N., Nonlinear dynamics of brake squeal, *Finite Elements in Analysis and Design*, 28, 69–82, 1997.
159. Flint, J., Instabilities in brake systems, Technical Report, SAE, Paper No. 920432, Warrendale, PA, 1992.
160. Hultén, J.O., Some drum brake squeal mechanisms, Technical Report, SAE, Paper No. 951280, Warrendale, PA, 1995.
161. Qiao, S. and Ibrahim, R.A., Stochastic dynamics of systems with friction-induced vibration, *Journal of Sound and Vibration*, 223(1), 115–140, 1999.
162. Sherif, H.A., Geometric induced instability in drum brakes, Technical Report, SAE, Paper No. 933072, Warrendale, PA, 1993.
163. So, J. and Leissa, A.W., Three-dimensional vibrations of thick circular and annular plates, *Journal of Sound and Vibration*, 209(1), 15–41, 1998.
164. Sugiyama, Y., Langthjem, M.A., and Ryu, B.-J., Realistic follower forces, *Journal of Sound and Vibration*, 225(4), 779–782, 1999.
165. Baillet, L., D'Errico, S., and Laulagnet, B., Understanding the occurrence of squealing noise using the temporal finite element method, *Journal of Sound and Vibration*, 292, 443–460, 2006.
166. Matozo, L.T., Soares, M.R.F., and Al-Qureshi, H.A., The effect of environmental humidity and temperature on friction level and squeal noise propensity for disc brake friction materials, Technical Report, SAE, Paper No. 2008-01-2534, Warrendale, PA, 2008.
167. Chen, G.X. and Zhou, Z.R., Correlation of a negative friction–velocity slope with squeal generation under reciprocating sliding conditions, *Wear*, 255, 376–384, 2003.
168. Hetzler, H. and Seemann, W., Brake vibrations at very low driving velocities, *Proceedings in Applied Mathematics and Mechanics*, 4, 87–88, 2004.

169. Hochlenert, D., Spelsberg-Korspeter, G., and Hagedorn, P., Friction induced vibrations in moving continua and their application to brake squeal, *Transactions of the ASME Journal of Applied Mechanics*, 74, 542–549, 2007.
170. Hoffmann, N. and Gaul, L., A sufficient criterion for the onset of sprag-slip oscillations, *Archive of Applied Mechanics*, 73, 650–660, 2004.
171. Hoffmann, N., Transient growth and stick-slip in sliding friction, *Transactions of the ASME*, 73, 643–647, 2006.
172. Pilipchuk, V.N. and Tan, C.A., Creep–slip capture as a possible source of squeal during decelerated sliding, *Nonlinear Dynamics*, 35, 259–285, 2004.
173. Lee, J.M., Yoo, S.W., and Kim, J.H., A study on the squeal of a drum brake which has shoes of non-uniform cross-section, *Journal of Sound and Vibration*, 240(5), 789–808, 2001.
174. Paliwal, M., Mahajan, A., Don, J., Chu, T., and Filip, P., Investigation of high-frequency squeal in a disc brake system using a friction layer-based coupling stiffness, *Proceedings of the IMechE, Part C*, 219, 513–522, 2005.
175. Park, J.-P. and Choi, Y.-S., Brake squeal noise due to disc run-out, *Proceedings of the IMechE*, *Part D*, 221, 811–821, 2007.
176. Sherif, H.A., Investigation on effect of surface topography of pad/disc assembly on squeal generation, *Wear*, 257, 687–695, 2004.
177. Sinou, J.-J., Thouverez, F., and Jezequel, L., Analysis of friction and instability by the centre manifold theory for a non-linear sprag-slip model, *Journal of Sound and Vibration*, 265, 527–559, 2003.
178. van der Auweraer, H., Hendricx, W., Vandenplas, B., and Pezzutto, A., Vibro-acoustic and structural dynamics analysis of brake noise, C603/019/2002, *IMechE* 2002.
179. Rhee, S. K., Jacko, M.G., and Tsang, P.H.S., The role of friction film in friction, wear, and noise of automotive brakes, Technical Report, SAE, Paper No. 900004, Warrendale, PA, 1990.
180. Bloch, A., Krishnaprasad, P.S., Marsden, J.E., and Ratiu, T.S., *Dissipation Induced Instabilities*, *Annales Institut H. Poincaré: Analyse Non Linéaire*, 11(1), 37–90, 1994.
181. Heppes, P., Noise insulators for brake squeal reduction—Influence and selection of the damping material, *The 2nd International Seminar on Automotive Braking*, D. Barton and M. Haigh (eds.), Leeds, U.K., pp. 15–26, 1998.
182. Tirovic, M. and Day, A.J., Disc brake interface pressure distributions, *Proceedings of the Institution of Mechanical Engineers*, 205, 137–146, 1991.
183. Day, A.J., Drum brake interface pressure distributions, *Proceedings of the Institution of Mechanical Engineers*, 205, 127–136, 1991.
184. Day, A.J., Tirovic, M., and Newcomb, T.P., Thermal effects and pressure distributions in brakes, *Proceedings of the Institution of Mechanical Engineers*, 205, 199–205, 1991.
185. Fieldhouse, J.D., A proposal to predict the noise frequency of a disc brake based on the friction pair interface geometry, *17th Annual SAE Brake Colloquium and Engineering Display*, Miami, FL, October 10–13, 1999, Technical Report, SAE, Paper No. 1999–01–3403, Warrendale, PA, 1999.
186. Kang, B. and Tan, C.A., Parametric instability of a Leipholz column under periodic excitation, *Journal of Sound and Vibration*, 229(5), 1097–1113, 2000.
187. Kobayashi, M. and Odani, N., Study on stabilisation friction coefficient of disc brake pads in cold condition, Technical Report, SAE, Paper No. 973030, Warrendale, PA, 1997.
188. Odani, N., Kobayashi, M., and Kakihara, K., Effects of transferred surface film on μ behaviour of disc brake pad in humidity environment, SAE, Paper No. 1999-01-3391, Warrendale, PA, 1999.
189. Lee, K. and Barber, J.R., Frictionally-excited thermoelastic instability in automotive disk brakes, *ASME Journal of Tribology*, 115, 607–614, 1993.
190. Lee, K. and Barber, J.R., The effect of shear tractions on frictionally-excited thermoelastic instability, *Wear*, 140, 237–242, 1993.
191. Mottershead, J.E., Ouyang, H., Cartmell, M.P., and Friswell, M.I., Parametric resonances in an annular disc, with a rotating system of distributed mass and elasticity; and the effects of friction and damping, *Proceedings of the Royal Society of London A*, 453, 1–19, 1997.
192. Ouyang, H. and Mottershead, J.E., Friction-induced parametric resonances in disc: Effect of a negative friction-velocity relationship, *Journal of Sound and Vibration*, 209(2), 251–264, 1998.
193. Mottershead, J.E., Ouyang, H., Cartmell, M.P., and Brookfield, D.J., Friction-induced vibration of an elastic slider on a vibrating disc, *International Journal of Mechanical Science*, 41, 325–336, 1999.
194. Thomsen, J.J., Using fast vibrations to quench friction-induced oscillations, *Journal of Sound and Vibration*, 228(5), 1079–1102, 1999.

195. Adams, G.G., Self excited oscillations of two elastic half spaces sliding with a constant friction coefficient, *ASME Journal of Applied Mechanics*, 62, 867–872, 1995.
196. Abu Bakar, A.R., Abdul Hamid, M.K., and Ouyang, H., A new prediction methodology for dynamic contact pressure distribution in a disc brake, *Jurnal Teknologi*, 45(A), 1–11, 2006.
197. Abu Bakar, A.R., Abdul Hamid, M.K., Dzakaria, A., Abd Ghani, B., and Mohamad, M., Stability analysis of disc brake squeal considering temperature effect, *Jurnal Mekanikal*, 22, 26–38, 2006.
198. Abu Bakar, A.R., Ouyang, H., James, S., and Li, S., Finite element analysis of wear and its effect on squeal, *Proceedings of the IMechE, Part D*, 222, 1153–1165, 2008.
199. Abu Bakar, A.R. and Ouyang, H., Wear prediction of friction material and brake squeal using the finite element method, *Wear*, 264, 1069–1076, 2008.
200. Afferrante, L. and Ciavarella, M., The thermoelastic Aldo contact model with frictional heating, *Journal of the Mechanics and Physics of Solids*, 52, 617–640, 2004.
201. Afferrante, L. and Ciavarella, M., "Frictionless" and "frictional" thermoelastic dynamic instability (TEDI) of sliding contacts, *Journal of the Mechanics and Physics of Solids*, 54(11), 2330–2353, 2006.
202. Afferrante, L., Ciavarella, M., and Barber, J.R., Sliding thermoelastodynamic instability, *Proceedings of the Royal Society A*, 462, 2161–2176, 2006.
203. Afferrante, L. and Ciavarella, M., A note on thermoelastodynamic instability (TEDI) for a 1D elastic layer: Force control, *International Journal of Solids and Structures*, 44, 1380–1390, 2007.
204. Afferrante, L. and Ciavarella, M., Thermo-elastic dynamic instability (TEDI) in frictional sliding of two elastic half-spaces, *Journal of the Mechanics and Physics of Solids*, 55, 744–764, 2007.
205. Baillet, L., D'Errico, S., and Berthier, Y., Influence of sliding contact local dynamics on macroscopic friction coefficient variation, *Revue européenne des éléments finis*, 10(10), 1–17, 2005.
206. Meziane, A., D'Errico, S., Baillet, L., and Laulagnet, B., Instabilities generated by friction in a pad–disc system during the braking process, *Tribology International*, 40, 1127–1136, 2007.
207. Massi, F., Berthier, Y., and Baillet, L., Contact surface topography and system dynamics of brake squeal, *Wear*, 265, 1784–1792, 2008.
208. Berger, E.J., Krousgrill, C.M., and Sadeghi, F., Friction-induced sliding instability in a multi-degree-of-freedom system with oscillatory normal forces, *Journal of Sound and Vibration*, 266, 369–387, 2003.
209. Zhu, B., Barton, D., and Brooks, P., Effects of thermal deformations on the squeal propensity of a simple automotive disc brake system, Technical Report, SAE, Paper No. 2008-01-2532, Warrendale, PA, 2008.
210. Lu, Y., Wang, K., and Jiao, L., Ceramic abrasives in brake friction materials, Technical Report, SAE, Paper No. 2008-01-2538, Warrendale, PA, 2008.
211. Imai, J., An acrylic rubber modified phenolic resin which has good flexibility at low temperature, Technical Report, SAE, Paper No. 2008-01-2539, Warrendale, PA, 2008.
212. Sanitate, F. and Schmitt, O., An investigation of metal pick-up generation on passenger car brake pads in correlation with deep rotor scoring, Technical Report, SAE, Paper No. 2008-01-2540, Warrendale, PA, 2008.
213. Nukumizu, K., Kobayashi, T., Abe, T., and Unno, M., Study of the formulation mechanism for metal pick-up on the frictional surface of a disc brake pad, Technical Report, SAE, Paper No. 2008-01-2541, Warrendale, PA, 2008.
214. Hassan, M.Z., Brooks, P.C., and Barton, D.C., Thermo-mechanical contact analysis of car disc brake squeal, Technical Report, SAE, Paper No. 2008-01-2566, Warrendale, PA, 2008.
215. Lou, G., Peters, B., Robere, M., Severnak, S., Spurr, W., and Gaulrapp, M., Lining life prediction based on rotor 1-D thermal model and applicability of wear versus temperature data, Technical Report, SAE, Paper No. 2008-01-2567, Warrendale, PA, 2008.
216. Steege, R. and Marx, F., A new approach to material compressibility of brake pads, Technical Report, SAE, Paper No. 2008-01-2572, Warrendale, PA, 2008.
217. Sanders, P.G., Dalka, T., and Hartsock, D., Friction material compressibility as a function of pressure, temperature, and frequency, Technical Report, SAE, Paper No. 2008-01-2574, Warrendale, PA, 2008.
218. Desplanques, Y. and Degallaix, G., Interaction between third-body flows and localisation phenomena during railway high-energy stop braking, Technical Report, SAE, Paper No. 2008-01-2583, Warrendale, PA, 2008.
219. Lee, Y.S., Brooks, P.C., Barton, D.C., and Crolla, D.A., A predictive tool to evaluate disc brake squeal propensity Part 1: The model philosophy and the contact problem, *International Journal of Vehicle Design*, 31(3), 289–308, 2003.
220. Cho, H.-J. and Cho, C.-D., A study of thermal and mechanical behaviour for the optimal design of automotive disc brakes, *Proceedings of the IMechE, Part D: Journal of Automobile Engineering*, 222, 895–915, 2008.

221. Cho, M.H., Cho, K.H., Kim, S.J., Kim, D.H., and Jang, H., The role of transfer layers on friction characteristics in the sliding interface between friction materials against gray iron brake disks, *Tribology Letters*, 20(2), 101–108, 2005.
222. Jibiki, T., Shima, M., Akita, H., and Tamura, M., A basic study of friction noise caused by fretting, *Wear*, 251, 1492–1503, 2001.
223. Kang, B. and Tan, C.A., Parametric instability of a Leipholz beam due to distributed frictional axial load, *International Journal of Mechanical Sciences*, 46, 807–825, 2004.
224. Kang, B. and Tan, C.A., Nonlinear response of a beam under distributed moving contact load, *Communications in Nonlinear Science and Numerical Simulation*, 11, 203–232, 2006.
225. Kang, J., Krousgrill, C.M., and Sadeghi, F., Dynamic instability of a thin circular plate with friction interface and its application to disc brake squeal, *Journal of Sound and Vibration*, 316, 164–179, 2008.
226. Kirillov, O.N., Subcritical flutter in the acoustics of friction, *Proceedings of the Royal Society A*, 464, 2321–2339, 2008.
227. Masoomi, M., Katbab, A.A., and Nazockdast, H., Damping behavior of the phenolic based composite friction materials containing thermoplastic elastomers (TPEs), *Iranian Journal of Chemistry and Chemical Engineering*, 25(3), 35–40, 2006.
228. Masoomi, M., Katbab, A.A., and Nazockdast, H., Noise generation and propagation control in disc brake systems using composite friction materials composed of thermoplastic elastomers as viscoelastic materials, *Polymer Composites*, 27, 461–469, 2006.
229. Moirot, F., Nguyen, Q.-S., and Oueslati, A., An example of stick–slip and stick–slip–separation waves, *European Journal of Mechanics A/Solids*, 22, 107–118, 2002.
230. Nguyen, Q.-S., *Instability and fiction*, *Comptes Rendus Mecanique*, 331, 99–112, 2003.
231. Lorang, X. and Nguyen, Q.S., Vibrations non-linéaires des structures en situation de contact frottant: Application aux crissements des freins à disque, *18 ème Congrès Français de Mécanique*, Grenoble, France, pp. 1–5, 2007.
232. Nguyen, Q.-S., Oueslati, A., Steindl, A., Teufel, A., and Troger, H., Travelling interface waves in a brake-like system under unilateral contact and Coulomb friction, *Comptes Rendus Mecanique*, 336, 203–209, 2008.
233. Österle, W., Griepentrog, M., Gross, Th., and Urban, I., Chemical and microstructural changes induced by friction and wear of brakes, *Wear*, 251, 1469–1476, 2001.
234. Ostermeyer, G.P., Friction and wear of brake systems, *Forschung im Ingenieurwesen*, 66, 267–272, 2001.
235. Ostermeyer, G.P., On the dynamics of the friction coefficient, *Wear*, 254, 852–858, 2003.
236. Mueller, M. and Ostermeyer, G.P., The interaction of surface topography and friction dynamics in brake systems described by a Cellular Automaton, *Proceedings in Applied Mathematics and Mechanics*, 5, 113–114, 2005.
237. Ostermeyer, G.P. and Müller, M., Dynamic interaction of friction and surface topography in brake systems, *Tribology International*, 39, 370–380, 2006.
238. Müller, M. and Ostermeyer, G.P., A Cellular Automaton model to describe the three-dimensional friction and wear mechanism of brake systems, *Wear*, 263, 1175–1188, 2007.
239. Müller, M. and Ostermeyer, G.P., Cellular automata method for macroscopic surface and friction dynamics in brake systems, *Tribology International*, 40, 942–952, 2007.
240. Ouyang, H. and Mottershead, J.E., Unstable travelling waves in the friction-induced vibration of discs, *Journal of Sound and Vibration*, 248(4), 768–779, 2001.
241. Rudolph, M. and Popp, K., The influence of contact properties on friction-induced brake vibrations, in: *Contact Mechanics* (eds. J.A.C. Martins and Manuel D.P. Monterio Marques), Kluwer Academic Publishers, the Netherlands, pp. 125–132, 2002.
242. Rudolph, M., Popp, K., and Hogenkamp, W., Computation of the temperature between brake disk and pad, *Proceedings in Applied Mathematics and Mechanics*, 3, 124–125, 2003.
243. Lindner, M., Sextro, W., and Popp, K., Hysteretic friction of a sliding rubber element, *Proceedings in Applied Mathematics and Mechanics*, 4, 101–102, 2004.
244. Sinou, J.-J., Fritz, G., and Jézéquel, L., The role of damping and definition of the robust damping factor for a self-exciting mechanism with constant friction, *Journal of Vibration and Acoustics*, 129, 297–306, 2007.
245. Stoimenov, B.L., Maruyama, S., Adachi, K., and Kato, K., The roughness effect on the frequency of frictional sound, *Tribology International*, 40, 659–664, 2007.
246. Duffour, P. and Woodhouse, J., Frictional instability of systems with a sliding point contact, C605/027/2002, *IMechE* 2002.
247. Brooks, P.C. and Sharp, R.S., A computational procedure based on eigenvalue sensitivity theory applicable to linear system design, *Journal of Sound and Vibration*, 114(1), 13–18, 1987.

248. D'Souza, A.F. and Dweib, A.H., Self-excited vibration induced by dry friction, part 1: Experimental study, *Journal of Sound and Vibration*, 137(2), 163–175, 1990.
249. Baba, H., Okade, M., and Takeuchi, T., Study on reducing low frequency brake squeal—From modal analysis of mounting bracket, Technical Report, SAE, Paper No. 952697, Warrendale, PA, 1995.
250. Ono, K., Chen, J.-S., and Bogy, D.B., Stability Analysis for the Head-Disk Interface in a Flexible Disk Drive, *ASME Journal of Applied Mechanics*, 58, 1005–1014, 1991.
251. Kung, S.-W., Dunlap, K.B., and Ballinger, R.S., Complex eigenvalue analysis for reducing low frequency brake squeal, Technical Report, SAE, Paper No. 2000-01-0444, Warrendale, PA, 2000.
252. Lee, C.W. and Kim, M.E., Separation and identification of travelling wave modes in rotating disk via directional spectral analysis, *Journal of Sound and Vibration*, 187(5), 851–864, 1995.
253. Lee, D. and Waas, A.M., Stability analysis of a rotating multi-layer annular plate with a stationary frictional follower load, *International Journal of Mechanical Sciences*, 39(10), 1117–l138, 1997.
254. Hinrichs, N., Oestreich, M., and Popp, K., Dynamics of oscillators with impact and friction, *Chaos, Solitons & Fractals*, 8(4), 535–558, 1997.
255. Hinrichs, N., Oestreich, M., and Popp, K., On the modelling of friction oscillators, *Journal of Sound and Vibration*, 216(3), 435–459, 1998.
256. Jezequel, L. and Lamarque, C.H., Analysis of non-linear dynamical systems by the normal form theory, *Journal of Sound and Vibration*, 149(3), 429–459, 1991.
257. Tseng, J.-G. and Wickert, G.A., On the vibration of bolted plate and flange assemblies, *Transactions of the ASME, Journal of Vibration and Acoustics*, 116, 468–473, 1994.
258. Tseng, J.-G. and Wickert, G.A., Vibration of an eccentrically clamped annular plate, *Transactions of the ASME, Journal of Vibration and Acoustics*, 116, 155–160, 1994.
259. Tseng, J.-G. and Wickert, G.A., Nonconservative stability of a friction loaded disk, *Transactions of the ASME, Journal of Vibration and Acoustics*, 120, 922–929, 1998.
260. Tseng, J.-G. and Wickert, G.A., Split vibration modes in acoustically-coupled disk stacks, *Transactions of the ASME, Journal of Vibration and Acoustics*, 120, 234–239, 1998.
261. Kim, M., Moon, J., and Wickert, J.A., Spatial modulation of repeated vibration modes in rotationally periodic structures, *Transactions of the ASME, Journal of Vibration and Acoustics*, 122, 62–68, 2000.
262. Afferrante, L., Ciavarella, M., Decuzzi, P., and Demelio, G., Transient analysis of frictionally excited thermoelastic instability in multi-disk clutches and brakes, *Wear*, 254, 136–146, 2003.
263. Akay, A., Acoustics of friction, *Journal of Acoustical Society of America*, 111(4), 1525–1548, 2002.
264. Andreaus, U. and Casini, P., Dynamics of friction oscillators excited by a moving base and/or driving force, *Journal of Sound and Vibration*, 245(4), 685–699, 2001.
265. Wu, G. and Zhou, G., Numerical investigation on brake noise mechanism incorporating nonlinear effects and complex eigenvalue extraction, Technical Report, SAE, Paper No. 2008-01-2535, Warrendale, PA, 2008.
266. Hoffrichter, W., Schmitt, O., and Fecske, L., Development of shim specifications, Technical Report, SAE, Paper No. 2008-01-2545, Warrendale, PA, 2008.
267. Bono, R.W. and Stultz, G.R., Resonant inspection applied to 100% testing of nodularity of cast ductile iron, Technical Report, SAE, Paper No. 2008-01-2577, Warrendale, PA, 2008.
268. Lee, Y.S., Brooks, P.C., Barton, D.C., and Crolla, D.A., A predictive tool to evaluate disc brake squeal propensity Part 2: System linearisation and modal analysis, *International Journal of Vehicle Design*, 31(3), 309–329, 2003.
269. Chen, G.X. and Zhou, Z.R., A self-excited vibration model based on special elastic vibration modes of friction systems and time delays between the normal and friction forces: A new mechanism for squealing noise, *Wear*, 262, 1123–1139, 2007.
270. Chen, G.X. and Zhou, Z.R., Time–frequency analysis of friction-induced vibration under reciprocating sliding conditions, *Wear*, 262, 1–10, 2007.
271. Chen, G.X., Liu, Q.Y., Jin, X.S., and Zhou, Z.R., Stability analysis of a squealing vibration model with time delay, *Journal of Sound and Vibration*, 311, 516–536, 2008.
272. Zhang, F., Cheng, L., Yam, L.H., and Zhou, L.M., Modal characteristics of a simplified brake rotor model using semi-analytical Rayleigh–Ritz method, *Journal of Sound and Vibration*, 297, 72–88, 2006.
273. Heilig, J. and Wauer, J., Stability of a nonlinear brake system at high operating speeds, *Nonlinear Dynamics*, 34, 235–247, 2003.
274. Jearsiripongkul, T., Chakraborty, G., and Hagedorn, P., Stability analysis of a floating caliper disc brake, *Proceedings in Applied Mathematics and Mechanics*, 2, 66–67, 2003.
275. Hoffmann, N. and Gaul, L., Effects of damping on mode-coupling instability in friction induced oscillations, *Zeitschrift für Angewandte Mathematik und Mechanik*, 83(8), 524–534, 2003.

276. Wagner, N. and Gaul, L., Eigenpath analyses of friction induced vibrations depending on the friction coefficient, *Proceedings in Applied Mathematics and Mechanics*, 3, 130–131, 2003.
277. Hoffmann, N., Bieser, S., and Gaul, L., Harmonic balance and averaging techniques for stick-slip limit-cycle determination in mode-coupling friction self-excited systems, *Technische Mechanik*, 24(3–4), 185–197, 2004.
278. Hoffmann, N. and Gaul, L., Non-conservative beating in sliding friction affected systems: Transient amplification of vibrational energy and a technique to determine optimal initial conditions, *Mechanical Systems and Signal Processing*, 18, 611–623, 2004.
279. Hoffmann, N., Wagner, N., and Gaul, L., Quenching mode-coupling friction-induced instability using high-frequency dither, *Journal of Sound and Vibration*, 279, 471–480, 2005.
280. Pilipchuk, V.N., Ibrahim, R.A., and Blaschke, P.G., Disc brake ring-element modeling involving friction-induced vibration, *Journal of Vibration and Control*, 8, 1085–1104, 2002.
281. Jerrelind, J. and Stensson, A., Nonlinear dynamics of parts in engineering systems, *Chaos, Solitons & Fractals*, 11, 2413–2428, 2000.
282. Joulin, O. and Picot, P., Analyse modale numérique d'un système de frein à disque, *Mécanique & Industries*, 4, 65–70, 2003.
283. Kang, J., Parametric study on friction-induced coupled oscillator, *Proceedings of the IMechE, Part C*, 222, 1381–1387, 2008.
284. Kessler, P., O'Reilly, O.M., Raphael, A.-L., and Zworski, M., On dissipation-induced destabilization and brake squeal: A perspective using structured pseudospectra, *Journal of Sound and Vibration*, 308, 1–11, 2007.
285. Kirillov, O.N., Gyroscopic stabilization of non-conservative systems, *Physics Letters A*, 359, 204–210, 2006.
286. Kirillov, O.N., Bifurcation of the roots of the characteristic polynomial and the destabilization paradox in friction induced oscillations, *Theoretical and Applied Mechanics*, 34(2), 87–109, 2007.
287. Kirillov, O.N., Destabilization paradox due to breaking the Hamiltonian and reversible symmetry, *International Journal of Non-Linear Mechanics*, 42, 71–87, 2007.
288. Kirillov, O.N., Stabilization and destabilization in non-conservative gyroscopic systems, *Proceedings in Applied Mathematics and Mechanics*, 7, 4050001–4050002, 2007.
289. Kliem, W., Mailybaev, A.A., and Pommer, C., Conditions revisited for asymptotic stability of pervasive damped linear systems, *Journal of Sound and Vibration*, 298, 471–474, 2006.
290. Li, Y. and Feng, Z.C., Bifurcation and chaos in friction-induced vibration, *Communications in Nonlinear Science and Numerical Simulation*, 9, 633–647, 2004.
291. Liu, P., Zheng, H., Cai, C., Wang, Y.Y., Lu, C., Ang, K.H., and Liu, G.R., Analysis of disc brake squeal using the complex eigenvalue method, *Applied Acoustics*, 68, 603–615, 2007.
292. Majcherczak, D. and Dufrénoy, P., Dynamic analysis of a disc brake under frictional and thermomechanical internal loading, *Archive of Applied Mechanics*, 75, 497–512, 2006.
293. Ouyang, H. and Mottershead, J.E., Dynamic instability of an elastic disk under the action of a rotating friction couple, *Journal of Applied Mechanics*, 71, 753–758, 2004.
294. Rahman, S., Stochastic dynamic systems with complex-valued eigensolutions, *International Journal for Numerical Methods in Engineering*, 71, 963–986, 2007.
295. Raman, A. and Mote Jr., C.D., Non-linear oscillations of circular plates near a critical speed resonance, *International Journal of Non-Linear Mechanics*, 34, 139–157, 1999.
296. Rusli, M. and Okuma, M., Effect of surface topography on mode-coupling model of dry contact sliding systems, *Journal of Sound and Vibration*, 308, 721–734, 2007.
297. Sinou, J.-J., Thouverez, F., and Jezequel, L., Center manifold and multivariable approximants applied to non-linear stability analysis, *International Journal of Non-Linear Mechanics*, 38(9), 1421–1442, 2003.
298. Sinou, J.-J., Thouverez, F., and Jezequel, L., Extension of the center manifold approach, using rational fractional approximants, applied to non-linear stability analysis, *Nonlinear Dynamics*, 33, 267–282, 2003.
299. Sinou, J.-J., Thouverez, F., and Jezequel, L., Application of a nonlinear modal instability approach to brake systems, *Journal of Vibration and Acoustics*, 126, 101–107, 2004.
300. Sinou, J.-J., Thouverez, F., and Jezequel, L., Non-linear stability analysis of a complex rotor/stator contact system, *Journal of Sound and Vibration*, 278, 1095–1129, 2004.
301. Sinou, J.-J., Thouverez, F., Jezeque, L., Dereure, O., and Mazet, G.-B., Friction induced vibration for an aircraft brake system—Part 2: Non-linear dynamics, *International Journal of Mechanical Sciences*, 48, 555–567, 2006.

302. Lignon, S., Sinou, J-J., and Jézéquel, L., Stability analysis and μ-synthesis control of brake systems, *Journal of Sound and Vibration*, 298(4–5), 1073–1087, 2006.
303. Fritz, G., Sinou, J.-J., Duffal, J.-M., and Jézéquel, L., Investigation of the relationship between damping and mode-coupling patterns in case of brake squeal, *Journal of Sound and Vibration*, 307, 591–609, 2007.
304. Sinou, J.-J. and Jézéquel, L., Mode coupling instability in friction-induced vibrations and its dependency on system parameters including damping, *European Journal of Mechanics A/Solids*, 26, 106–122, 2007.
305. Lesaffre, N., Sinou, J.-J., and Thouverez, F., Stability analysis of rotating beams rubbing on an elastic circular structure, *Journal of Sound and Vibration*, 299, 1005–1032, 2007.
306. Sinou, J.-J. and Jézéquel, L., The influence of damping on the limit cycles for a self-exciting mechanism, *Journal of Sound and Vibration*, 304, 875–893, 2007.
307. Hervé, B., Sinou, J.-J., Mahé, H., and Jézéquel, L., Analysis of squeal noise and mode coupling instabilities including damping and gyroscopic effects, *European Journal of Mechanics A/Solids*, 27, 141–160, 2008.
308. Sinou, J.-J. and Jézéquel, L., On the stabilizing and destabilizing effects of damping in a non-conservative pin-disc system, *Acta Mechanica*, 199, 43–52, 2008.
309. Steindl, A., Bifurcations of Stick-Slip-Separation Waves in a Brake-like System, *Proceedings in Applied Mathematics and Mechanics*, 6, 337–338, 2006.
310. Steindl, A., Hopf bifurcations of travelling waves in a brake-like system, *Proceedings in Applied Mathematics and Mechanics*, 7, 4050025–4050026, 2007.
311. Chang, J.Y. and Wickert, J.A., Response of modulated doublet modes to travelling wave excitation, *Journal of Sound and Vibration*, 242(1), 69–83, 2001.
312. Young, T.H. and Lin, C.Y., Stability of a spinning disk under a stationary oscillating unit, *Journal of Sound and Vibration*, 298, 307–318, 2006.
313. Zhou, M., Wang, Y., and Huang, Q., Study on the stability of drum brake non-linear low frequency vibration model, *Archive of Applied Mechanics*, 77, 473–483, 2007.
314. Zhou, M., Chen, Y., Wang, Y., and Zhou, J., Study on the limit cycle oscillation of the drum brake nonlinear vibration model at low frequency, *International Journal of Mechanics and Materials in Design*, 4, 317–324, 2008.
315. Lee, H., Modal acoustic radiation characteristics of a thick annular disk, Dissertation of The Ohio State University, Columbus, OH, 2003.
316. Liles, G.D., Analysis of disc brake squeal using finite element methods, Technical Report, SAE, Paper No. 891150, Warrendale, PA, 1989.
317. Chen, J.-S. and Bogy, D.B., Effects of load parameters on the natural frequencies and stability of a flexible spinning disk with a stationary load system, *Transactions of the American Society of Mechanical Engineers, Journal of Applied Mechanics*, 59(2), 5230–5235, 1992.
318. Chen, J.-S. and Bogy, D.B., Mathematical structure of modal interactions in a spinning disk-stationary load system, *Transactions of the American Society of Mechanical Engineers, Journal of Applied Mechanics*, 59(2), 390–397, 1992.
319. Hartsock, D.L., Hecht, R.L., and Fash, J.W., Parametric analyses of thermoelastic instability in disc brakes, *International Journal of Vehicle Design*, 21(4/5), 510–526, 1999.
320. Du, S. and Fash, J.W., Finite element analysis of frictionally-excited thermoelastic instability in 3D annular disk, *International Journal of Vehicle Design*, 23(3/4), 203–217, 2000.
321. McMillan, A.J., A non-linear friction model for self-excited vibrations, *Journal of Sound and Vibration*, 205(3), 323–335, 1997.
322. Nagy, L.I., Cheng, J., and Hu, Y.-K., A new method development to predict brake squeal occurrence, Technical Report, SAE, Paper No. 942258, Warrendale, PA, 1994.
323. Hu, Y.-K. and Nagy, L.I., Brake squeal analysis by using nonlinear transient finite element method, Technical Report, SAE, Paper No. 971510, Warrendale, PA, 1997.
324. Hohmann, C., Schiffner, K., Oerter, K., and Reese, H., Contact analysis for drum brakes and disk brakes using ADINA, *Computers and Structures*, 72, 185–198, 1999.
325. Ahmed, I.L.M., Leung, P.S., and Datta, P.K., Contact analysis of the disc brake by finite element methods, C605/001/2002, *IMechE* 2002.
326. Berger, E.J., Friction modeling for dynamic system simulation, *ASME Applied Mechanics Reviews*, 55(6), 535–577, 2002.
327. Gobbi, M., Guarneri, P., and Mastinu, G., Optimal robust design optimisation with application to a piezoelectric brake, Technical Report, SAE, Paper No. 2008-01-2554, Warrendale, PA, 2008.

328. Choudhary, A. and Stringham Jr., W.J., Improvements in simulation techniques to better understand disc brake dynamics, Technical Report, SAE, Paper No. 2008-01-2564, Warrendale, PA, 2008.
329. Söderberg, A., Sellgren, U., and Andersson, S., Using finite element analysis to predict the brake pressure needed for effective rotor cleaning in disc brakes, Technical Report, SAE, Paper No. 2008-01-2565, Warrendale, PA, 2008.
330. Aleksendric, D., Duboka, C., and Cirovic, V., Intelligent control of disc brake operation, Technical Report, SAE, Paper No. 2008-01-2570, Warrendale, PA, 2008.
331. Österle, W., Kloß, H., and Dmitriev, A.I., Modeling of friction evolution and assessment of impacts on vibration excitation at the pad–disc interface, Technical Report, SAE, Paper No. 2008-01-2582, Warrendale, PA, 2008.
332. Kang, H.I., Brooks, P.C., and Barton, D.C., Drum brake squeal prediction using a parametric finite element model, C605/007/2002, *IMechE*, London, U.K., 2002.
333. Lee, Y.S., Brooks, P.C., Barton, D.C., and Crolla, D.A., A predictive tool to evaluate disc brake squeal propensity Part 3: Parametric design studies, *International Journal of Vehicle Design*, 31(3), 330–353, 2003.
334. Chatterjee, S., Non-linear control of friction-induced self-excited vibration, *International Journal of Non-Linear Mechanics*, 42, 459–469, 2007.
335. Chatterjee, S., Time-delayed feedback control of friction-induced instability, *International Journal of Non-Linear Mechanics*, 42, 1127–1143, 2007.
336. Das, J. and Mallik, A.K., Control of friction driven oscillation by time-delayed state feedback, *Journal of Sound and Vibration*, 297, 578–594, 2006.
337. Emira, M.N.A., Friction-induced oscillations of a slider: Parametric study of some system parameters, *Journal of Sound and Vibration*, 300, 916–931, 2007.
338. Feng, Q. and Zhang, X.-T., The discrete models on a frictional single degree of freedom system, *Applied Mathematics and Mechanics*, 22(8), 956–964, 2001.
339. Feng, Q., A discrete model of a stochastic friction system, *Computer Methods in Applied Mechanics and Engineering*, 192, 2339–2354, 2003.
340. Guan, D., Su, X., and Zhang, F., Sensitivity analysis of brake squeal tendency to substructures' modal parameters, *Journal of Sound and Vibration*, 291, 72–80, 2006.
341. Heckmann, A., A brake model with thermoelastic disc for the analysis of vehicle judder vibrations, *Vehicle System Dynamics*, 44(1), 360–367, 2006.
342. Jearsiripongkul, T. and Hagedorn, P., Parameter estimation of the disk in a floating caliper disk brake model with respect to squeal, *Proceedings in Applied Mathematics and Mechanics*, 4, 93–94, 2004.
343. Jearsiripongkul, T., A simplified model of the floating caliper disk brake with respect to high frequency noise, *Thammasat International Journal of Science and Technology*, 11(2), 61–68, 2006.
344. Huang, J., Krousgrill, C.M., and Bajaj, A.K., Modeling of automotive drum brakes for squeal and parameter sensitivity analysis, *Journal of Sound and Vibration*, 289, 245–263, 2006.
345. Huang, J., Krousgrill, C.M., and Bajaj, A.K., An efficient approach to estimate critical value of friction coefficient in brake squeal analysis, *Transactions of the ASME*, 74, 534–541, 2007.
346. Kerschen, G., Worden, K., Vakakis, A.F., and Golinval, J.-C., Past, present and future of nonlinear system identification in structural dynamics, *Mechanical Systems and Signal Processing*, 20, 505–592, 2006.
347. Komzsik, L., Implicit computational solution of generalized quadratic, *Finite Elements in Analysis and Design*, 37, 799–810, 2001.
348. Lou, G., Wu, T.W., and Ba, Z., Disk brake squeal prediction using the ABLE algorithm, *Journal of Sound and Vibration*, 272, 731–748, 2004.
349. Ouyang, H., Mottershead, J.E., and Li, W., A moving-load model for disc-brake stability analysis, *ASME Journal of Vibration and Acoustics*, 125(1), 1–6, 2003.
350. Paliwal, M., Mahajan, A., Don, J., Chu, T., and Filip, P., Noise and vibration analysis of a disc–brake system using a stick–slip friction model involving coupling stiffness, *Journal of Sound and Vibration*, 282, 1273–1284, 2005.
351. Shin, K., Brennan, M.J., Oh, J.-E., and Harris, C.J., Analysis of disc brake noise using a two-degree-of-freedom model, *Journal of Sound and Vibration*, 254(5), 837–848, 2002.
352. Shin, K., Brennan, M.J., Joe, Y.-G., and Oh, J.-E., Simple models to investigate the effect of velocity dependent friction on the disc brake squeal noise, *International Journal of Automotive Technology*, 5(1), 61–67, 2004.
353. Oh, J.-E., Joe, Y.-G., and Shin, K., Analysis of out-of-plane motion of a disc brake system using a two-degree-of-freedom model with contact stiffness, *Proceedings of the IMechE, Part D*, 219, 869–879, 2005.

354. Joe, Y.-G., Cha, B.-G., Sim, H.-J., Lee, H.-J., and Oh, J.-E., Analysis of disc brake instability due to friction-induced vibration using a distributed parameter model, *International Journal of Automotive Technology*, 9(2), 161–171, 2008.
355. Fritz, G., Sinou, J.-J., Duffal, J.-M., and Jézéquel, L., Effects of damping on brake squeal coalescence patterns—Application on a finite element model, *Mechanics Research Communications*, 34, 181–190, 2007.
356. Coudeyras, N., Sinou, J.-J., and Nacivet, S., A new treatment for predicting the self-excited vibrations of nonlinear systems with frictional interfaces: The Constrained Harmonic Balance Method, with application to disc brake squeal, *Journal of Sound and Vibration*, 319, 1175–1199, 2009.
357. Slavič, J., Bryant, D., and Boltežara, M., A new approach to roughness-induced vibrations on a slider, *Journal of Sound and Vibration*, 306, 732–750, 2007.
358. Stoimenov, B.L. and Mukai, T., A model of adhesion excited micro-vibrations, *Tribology Letters*, 29, 129–137, 2008.
359. Thomsen, J.J. and Fidlin, A., Analytical approximations for stick–slip vibration amplitudes, *International Journal of Non-Linear Mechanics*, 38, 389–403, 2003.
360. Trichés Jr., M.J., Gerges, S.N.Y., and Jordan, R., Analysis of brake squeal noise using the finite element method: A parametric study, *Applied Acoustics*, 69, 147–162, 2008.
361. Duffour, P. and Woodhouse, J., Instability of systems with a frictional point contact. Part 1: Basic modelling, *Journal of Sound and Vibration*, 271, 365–390, 2004.
362. Duffour, P. and Woodhouse, J., Instability of systems with a frictional point contact. Part 2: Model extensions, *Journal of Sound and Vibration*, 271, 391–410, 2004.
363. Abendroth, H., Advances in brake NVH test equipment, *Automotive Engineering International*, 107(2), 60–63, 1999.
364. Fieldhouse, J.D. and Newcomb, T.P., The application of holographic interferometry to the study of disc brake noise, Technical Report, SAE, Paper No. 930805, Warrendale, PA, 1993.
365. Wang, S.K. and Woodhouse, J., A novel measurement of dynamic friction, Technical Report, SAE, Paper No. 2008-01-2536, Warrendale, PA, 2008.
366. Fieldhouse, J.D., Bryant, D., Crampton, A., Talbot, C., and Layfield, J., A study of thermal judeer on a laboratory dynamometer, Technical Report, SAE, Paper No. 2008-01-2542, Warrendale, PA, 2008.
367. Backstrom, A., An experimental investigation of brake rotor DTV under laboratory conditions, Technical Report, SAE, Paper No. 2008-01-243, Warrendale, PA, 2008.
368. Sardá, A., Haag, M., Winner, H., and Semsch, M., Experimental investigation of hot spots and thermal judder, Technical Report, SAE, Paper No. 2008-01-2544, Warrendale, PA, 2008.
369. Hwang, P., Wu, X., and Jeon, Y.B., Repeated brake temperature analysis of ventilated brake disc on the downhill road, Technical Report, SAE, Paper No. 2008-01-2571, Warrendale, PA, 2008.
370. Schroeder, E.J. and Stringham, W., An experimental methodology to determine the low energy wear model, Technical Report, SAE, Paper No. 2008-01-2573, Warrendale, PA, 2008.
371. Chen, G.X., Zhou, Z.R., Kapsa, P., and Vincent, L., Experimental investigation into squeal under reciprocating sliding, *Tribology International*, 36, 961–971, 2003.
372. Chen, G.X. and Zhou, Z.R., Experimental observation of the initiation process of friction-induced vibration under reciprocating sliding conditions, *Wear*, 259, 277–281, 2005.
373. Cunefare, K.A. and Graf, A.J., Experimental active control of automotive disc brake rotor squeal using dither, *Journal of Sound and Vibration*, 250(4), 579–590, 2002.
374. Fieldhouse, J.D., Low frequency drum brake noise investigation using a 1/4 vehicle test rig, Technical Report, SAE, Paper No. 2000-01-0448, Warrendale, PA, 2000.
375. Giannini, O., Akay, A., and Xu, Z., A laboratory brake for the study of automotive brake noise, *Proceedings of IMAC XX*, Paper No. 373, Los Angeles, CA, 2002.
376. Giannini, O. and Massi, F., An experimental study on the brake squeal noise, *Proceedings of ISMA*, Leuven, Belgium, pp. 3411–3426, 2004.
377. Massi, F., Giannini, O., and Baillet, L., Brake squeal as dynamic instability: An experimental investigation, *Journal of the Acoustical Society of America*, 120(3), 1388–1399, 2006.
378. Giannini, O. and Massi, F., Characterization of the high-frequency squeal on a laboratory brake setup, *Journal of Sound and Vibration*, 310, 394–408, 2008.
379. Schlagner, S. and von Wagner, U., Evaluation of automotive disk brake noise behavior using piezoceramic actuators and sensors, *Proceedings in Applied Mathematics and Mechanics*, 7, 4050031–4050032, 2007.
380. Rudolph, M. and Popp, K., Vibration control to avoid stick–slip motion, *Journal of Vibration and Control*, 10, 1585–1600, 2004.

381. Kröger, M., Neubauer, M., and Popp, K., Experimental investigation on the avoidance of self-excited vibrations, *Philosophical Transactions of the Royal Society A*, 366, 785–810, 2008.
382. Sergienko, V.P., Bukharov, S.N., and Kupreev, A.V., Noise and vibration in brake systems of vehicles. Part 1: Experimental procedures, *Journal of Friction and Wear*, 29(3), 234–241, 2008.
383. Duffour, P. and Woodhouse, J., Instability of systems with a frictional point contact—Part 3: Experimental tests, *Journal of Sound and Vibration*, 304(2007), 186–200, 2007.
384. Bergman, F., Eriksson, M., and Jacobson, S., A software-based measurement system for testing and analysis of automotive brake squeal, *Tribotest Journal*, 5(3), 265–275, 1999.
385. Reeves, M., Taylor, N., Edwards, C., Williams, D., and Buckberry, C.H., A study of brake disc modal behaviour during squeal generation using high-speed electronic speckle pattern interferometry and near-speed sound pressure measurements, *Proceedings of the Institution of Mechanical Engineers, Part D*, 214, 285–296, 2000.
386. Okamura, T., Yumoto, H., and Imasaki, M., Factorial effect analysis of material and dimensional properties of brake discs for reducing variation in natural frequencies, Technical Report, SAE, Paper No. 2008-01-2533, Warrendale, PA, 2008.
387. Barigozzi, G., Perdichizzi, A., and Donati, M., Combined experimental and CFD investigation of brake discs aero-thermal performances, Technical Report, SAE, Paper No. 2008-01-2550, Warrendale, PA, 2008.
388. Strom, P., Dynamic vehicle weight reduction and safety enhancement, Technical Report, SAE, Paper No. 2008-01-2551, Warrendale, PA, 2008.
389. Huang, C.-T., Chen, C.-T., Cheng, S.-Y., Chen, B.-R., and Huang, M.-H., Design and testing of a new electric parking brake actuator, Technical Report, SAE, Paper No. 2008-01-2555, Warrendale, PA, 2008.
390. Farhang, K., Ozcan, S., and Filip, P., The effect of wear groove on vibration and noise of aircraft brakes: Theorical and experimental evidence, Technical Report, SAE, Paper No. 2008-01-2557, Warrendale, PA, 2008.
391. Eriston, L.J. and Miles, M.D., Retrofittable regenerative braking in heavy vehicle applications, Technical Report, SAE, Paper No. 2008-01-2558, Warrendale, PA, 2008.
392. Surampudi, B., Redfield, J., and Ostrowski, G.J., Mild regenerative braking to enhance fuel economy via lowered engine load due to alternator, Technical Report, SAE, Paper No. 2008-01-2560, Warrendale, PA, 2008.
393. Ma, J., Liu, X., and Sun, J., A new method of inertia simulation in brake dynamometer testing, Technical Report, SAE, Paper No. 2008-01-2578, Warrendale, PA, 2008.
394. Fieldhouse, J.D. and Steel, W.P., A study of brake noise and the influence of the centre of pressure at the disc/pad interface, the coefficient of friction and calliper mounting geometry, *Proceedings of the Institution of Mechanical Engineers*, *Part D*, 217, 957–973, 2003.
395. Talbot, C. and Fieldhouse, J.D., Fourier analysis of holographic data from a noisy disc brake and its implication for modelling, *Proceedings of the Institution of Mechanical Engineers*, *Part D*, 217, 975–984, 2003.
396. Casini, P., Giannini, O., and Vestroni, F., Experimental evidence of non-standard bifurcations in non-smooth oscillator dynamics, *Nonlinear Dynamics*, 46, 259–272, 2006.
397. Guan, D. and Huang, J., The method of feed-in energy on disc brake squeal, *Journal of Sound and Vibration*, 261, 297–307, 2003.
398. Hetzler, H. and Seemann, W., Friction modes in low frequency disc-brake noise—Experimental results and implications on modelling, *Proceedings in Applied Mathematics and Mechanics*, 6, 307–308, 2006.
399. Hochlenert, D. and Hagedorn, P., Control of disc brake squeal—Modelling and experiments, *Structural Control and Health Monitoring*, 13, 260–276, 2006.
400. Qiao, S., Beloiu, D.M., and Ibrahim, R.A., Deterministic and stochastic characterization of friction-induced vibration in disc brakes, *Nonlinear Dynamics*, 36, 361–378, 2004.
401. Neubauer, M. and Oleskiewicz, R., Brake squeal control with shunted piezoceramics—Efficient modelling and experiments, *Proceedings of the IMechE, Part D*, 222, 1141–1151, 2008.
402. Sinou, J-J., Dereure, O., Mazet, G-B., Thouverez, F., and Jezequel, L., Friction induced vibration for an aircraft brake system—Part 1: Experimental approach and stability analysis, *International Journal of Mechanical Sciences*, 48(5), 536–554, 2006.
403. Butlin, T. and Woodhouse, J., Sensitivity of friction-induced vibration in idealised systems, *Journal of Sound and Vibration*, 319, 182–198, 2009.
404. DiLisio, P., Parisi, R., Rieker, J., and Stringham, W., Brake noise resolution on the 1998 Mercedes-Benz M-class, Technical Report, SAE, Paper No. 982245, Warrendale, PA, 1998.
405. Liu, W. and Pfeifer, J., Reducing high frequency disc brake squeal by pad shape optimization, Technical Report, SAE, Paper No. 2000-01-0447, Warrendale, PA, 2000.

406. Malosh, J.B., Disc brake noise reduction through metallurgical control of rotor resonances, Technical Report, SAE, Paper No. 982236, Warrendale, PA, 1998.
407. Doi, K., Mibe, T., Matsui, H., Tamasho, T., and Nakanishi, H., Brake judder reduction technology-brake design technique including friction material formulation, *JSAE Review*, 21, 497–502, 2000.
408. Nakajima, T. and Okada, Y., Study on reduction method of brake squeal, Technical Report, SAE, Paper No. 973029, Warrendale, PA, 1997.
409. Grochowicz, J., Experimentelle und theoretische Untersuchungen zu Bremsdruck-und Bremsmomentenschwankungen an Pkw-Scheibenbremsen, Dissertation, Dresden Technical University, Dresden, Germany, 1996.
410. Ganguly, S., Tong, H., and Karpenko, Y., A systems approach to eliminating squeal in a drum brake, Technical Report, SAE, Paper No. 2008-01-2531, Warrendale, PA, 2008.
411. Kappagantu, R.V. and Denys, E., Geometric tuning of insulators for brake squeal attenuation, Technical Report, SAE, Paper No. 2008-01-2546, Warrendale, PA, 2008.
412. Dickinson, R.C., Vehicle brake noise shims: A tutorial for continuous product improvements, Technical Report, SAE, Paper No. 2008-01-2547, Warrendale, PA, 2008.
413. Shevket, C., Zhou, X.B., and Kapaan, H., Sacrificial shield for the wheel hub bearing flange to maintain low brake rotor lateral runout, Technical Report, SAE, Paper No. 2008-01-2548, Warrendale, PA, 2008.
414. Kappagantu, R.V., Vibro-impact rotor dampers for brake squeal attenuation—Towards an insulator free design to quell squeal, Technical Report, SAE, Paper No. 2008-01-2549, Warrendale, PA, 2008.
415. Nishiwaki, M., Abe, K., Yanagihara, H., Stankovic, I., Nagasawa, Y., and Wakamatsu, S., A study on friction materials for brake squeal reduction by nanotechnology, Technical Report, SAE, Paper No. 2008-01-2581, Warrendale, PA, 2008.
416. Chatterjee, S., On the design criteria of dynamic vibration absorbers for controlling friction-induced oscillations, *Journal of Vibration and Control*, 14(3), 397–415, 2008.
417. Cunefare, K.A. and Graf, A.J., Disc brake rotor squeal suppression using dither control, Technical Report, SAE, Paper No. 2001-01-1605, Warrendale, PA, 2001.
418. Hammerström, L. and Jacobson, S., Surface modification of brake discs to reduce squeal problems, *Wear*, 261, 53–57, 2006.
419. Hwang, I.-J. and Park, G.-J., Mode and design sensitivity analyses for brake judder reduction, *Proceedings of the IMechE, Part D*, 222, 1259–1272, 2008.
420. Masoomi, M., Katbab, A.A., and Nazockdast, H., Reduction of noise from disc brake systems using composite friction materials containing thermoplastic elastomers (TPEs), *Applied Composite Materials*, 13, 305–319, 2006.
421. Masoomi, M., Katbab, A.A., and Nazockdast, H., Sound attenuation in composite friction materials containing thermoplastic elastomers, *Journal of Applied Polymer Science*, 101, 2187–2194, 2006.
422. Michaux, M.A., Ferri, A.A., and Cunefare, K.A., Effect of waveform on the effectiveness of tangential dither forces to cancel friction-induced oscillations, *Journal of Sound and Vibration*, 311, 802–823, 2008.
423. Neubauer, M. and Oleskiewicz, R., Suppression of brake squeal using shunted piezoceramics, *Journal of Vibration and Acoustics*, 130, 1–8, 2008.
424. Ouyang, H. and Mottershead, J.E., Optimal suppression of parametric vibration in discs under rotating frictional loads, *Proceedings of the Institution of Mechanical Engineers, Part C*, 215, 65–75, 2001.
425. Trichés Jr., M.J., Gerges, S.N.Y., and Jordan, R., Reduction of squeal noise from disc brake systems using constrained layer damping, *ABCM Journal of the Brazilian Society of Mechanical Sciences and Engineering*, XXVI(3), 340–348, 2004.
426. Weller, J.A., Roman traction systems, www.humanist.de/rome/rts 2013.
427. Abdo, J.A. and Meinhardt, G.A., Vibration response of a disc brake: Evaluation and design, *International Journal of Applied Mechanics and Engineering*, 11(3), 467–479, 2006.
428. Brommundt, E., Ein Reibschwinger mit Selbsterregung ohne fallende Reibkennlinie, *Zeitschrift fur Angewandte Mathematik und Mechanik*, 75(11), 811–820, 1995.
429. D'Souza, A.F. and Dweib, A.H., Self-excited vibration induced by dry friction, part 2: Stability and limit cycle analysis, *Journal of Sound and Vibration*, 137(2), 177–190, 1990.
430. Schlagner, S. and von Wagner, U., Quietschen von KFZ-Scheibenbremsen, *Proceedings in Applied Mathematics and Mechanics*, 6, 329–330, 2006.
431. Ostermeyer, G.P. and Müller, M., New insights into the tribology of brake systems, *Proceedings of the IMechE, Part D*, 222, 1167–1200, 2008.

25 Steering System

Ichiro Kageyama

CONTENTS

25.1 Introduction 919
25.2 Basic Configuration of Steering Equipment 919
25.2.1 Basic Configuration 920
25.3 History of Steering Equipment 920
25.3.1 History of Steering System Mechanisms 920
25.3.2 History of Power Steering 923
25.4 Basic Theory of Steering Systems 924
25.4.1 Ackerman Steering 924
25.4.2 Modeling of the Steering System 926
25.4.2.1 Equivalent Moment of Inertia in Steering Systems 926
25.4.2.2 Modeling for Simple Steering Systems 927
25.4.2.3 Detailed Model of the Steering System 929
25.5 Power Steering 930
25.5.1 Types of Power Steering 931
25.5.2 Modeling for Power Steering Systems 932
25.5.3 Example Analysis Including Power Steering System 934
25.5.4 Electric Power Steering 934
25.5.4.1 Modeling for EPS 935
25.5.4.2 Block Diagram of Control System 936
25.5.4.3 Vehicle Model Including EPS 937
References 938

25.1 INTRODUCTION

Steering equipment is configured to enable a person to control the direction of a vehicle. The centripetal force when a vehicle turns is due to friction between the tires and the road and is configured to control the yawing moment applied to the vehicle by controlling the direction of the friction force. To control this yawing moment, methods using both lateral and longitudinal forces need to be considered. The former is called the steering method and is adopted by vehicles mainly using tires. The latter is called the brake control method and is mainly used by crawler-type vehicles. Further, in recent years, some of the vehicles that use tires have adopted the use of traction and braking force systems called the yaw moment control method, but currently, in this chapter, it has been decided to limit descriptions to control methods using steering, with the objective of supplementary stability control. (For more details on yaw moment control, see Chapters 30, 32, and 39.)

25.2 BASIC CONFIGURATION OF STEERING EQUIPMENT

Control methods are methods of controlling the size and direction of the lateral force by controlling the tire direction and can be divided into types that control the front wheels (FWS, front-wheel steering) and types whose control also include the rear wheels (4WS, four-wheel steering or AWS, all-wheel steering). Usually, analysis of the 4WS system is carried out in the field of vehicle

handling and control. Moreover, in recent years, the steer-by-wire system controlled by computers is one of the new topics in the field of steering systems. For example, by adding the factor of steering angular velocity to the output of steer angle, the delay in the driver's steer can be compensated. The analysis of such systems is also carried out in the field of vehicle handling and control. This chapter, therefore, mainly describes the configuration of the front-wheel steering system, types of power-assisted steering systems, and equations of motion for steering systems and power-assist systems.

25.2.1 Basic Configuration

Steering equipment can generally be thought of as comprising three parts: the control mechanism, the gear mechanism, and the link mechanism. The control mechanism is mainly the part that enables the driver to steer directly and comprises the steering wheel, steering shaft, and column. The steering force created by the driver is transmitted to the gear and link mechanisms and also plays the role of transmitting to the driver the aligning moments created by the tires and the kick-back from the road surface.

The gear mechanism decelerates the steering shaft rotations and is also responsible for changing the direction of operations of the steering mechanism, as well as increasing the steering moment. Vehicles in recent years have used a configuration that changes the gear ratio according to the rotation angle. Further, the nomenclature of the steering system changes depending on the type and configuration of the gears used. In addition, the history of steering systems greatly depends on changes to the gear configuration.

The link mechanism is the part that transmits the movements of the gear mechanism to the front wheels, and its configuration elements depend on the type of gear mechanism, but the configuration elements include a pitman arm, relay rod, tie rod, and knuckle arm as shown in Figure 25.1.

Further, in recent years, rack-and-pinion-type gear mechanisms have been widely adopted, and the gear mechanism has been combined with some link mechanisms. Generally, when turning a vehicle at extremely low speeds, it is necessary to change the steering angle due to the difference between the turning radius of the right and left wheels, and this is achieved by the design of the link mechanism.

These steering elements—the steering shaft, the link mechanism, and the bush for the joints, in particular—have elasticity so that these allow a deformation angle between the steering wheel and the steering tires. By this steering elasticity, the impact received from the road through the tire, especially for longitudinal and rotational direction such as kickback, is eased. On the one hand, it also greatly affects the steering stability by trying to return the steering angle to the central position due to the aligning moment (self-aligning torque) created by the front wheels during turning. It is necessary to decrease the value of cornering stiffness for the steered tire at low degrees of freedom without considering the steering system. In addition, hysteresis is described in the steering system between the steering wheel and the steering angle of the front wheels due to backlash and friction between the gear parts and the joints, and, further, nonlinearity also occurs depending on the link configuration and variable gear ratios (VGRs); therefore, it is necessary to consider these according to the modeling objectives. A typical steering characteristic between the steering torque and steering angle is described Figure 25.2.

25.3 HISTORY OF STEERING EQUIPMENT

25.3.1 History of Steering System Mechanisms

In the early days of car building, which developed from horse-drawn carriages, there were mainly two configurations of steering systems. One was to adopt the steering mechanism of the horse-drawn carriage, which rigidly fused the left and right wheels, and the other was a three-wheeled vehicle that combined the steered wheels into one. The gun carriage invented by Nicolas-Joseph Cugnot

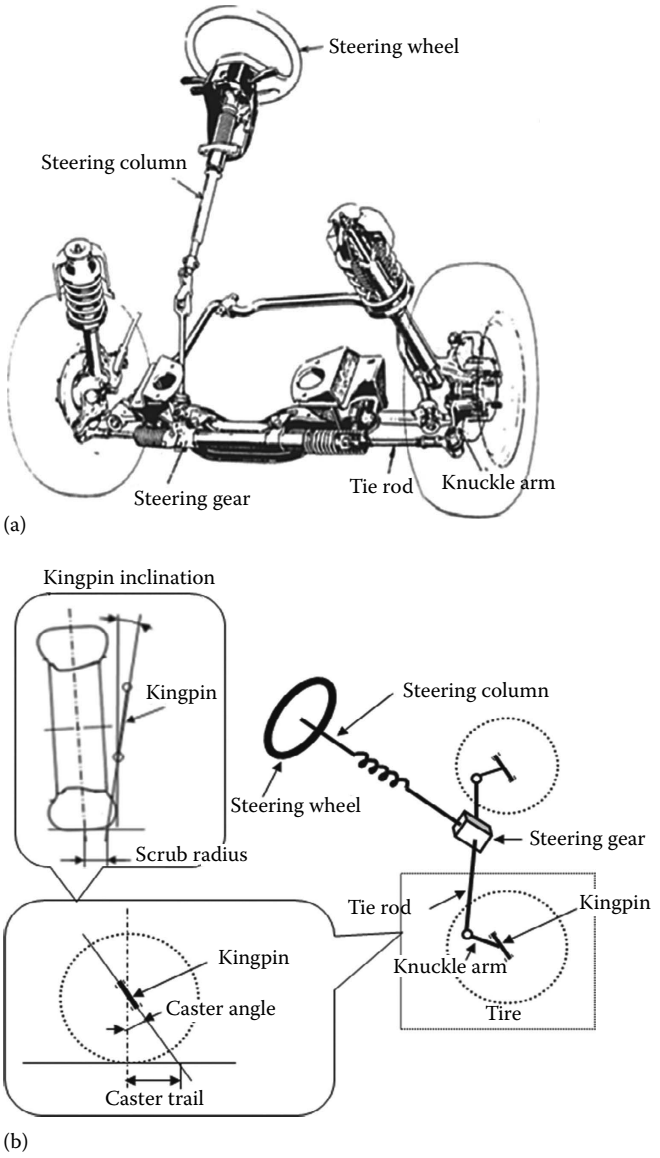


FIGURE 25.1 Outline of steering system. (From Komamura, S. et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.)

from France in 1769 [1,2], which can be considered to be the first motor vehicle, was a three-wheeler and the front wheel could be steered. And, in particular, the steering system was configured to rotate a part of the ring gear mounted directly on to the steering wheel by performing through two-step gears from the steering shaft as shown in Figure 25.3.

Further, 250 years earlier, in 1480, diagrams of automobiles drawn by Leonardo da Vinci [1], while not describing these steering mechanisms in detail, handled the steering wheels as a single wheel as shown in Figure 25.4.

When the geometry of turning is considered, these have the advantages that the wheels do not slide sideways, steering torque can be minimized, and the mechanism is comparatively simple. This three-wheel mechanism was also adopted by Karl Benz for his automobiles, which are one of the oldest gasoline-powered automobiles. When stability during turning is considered, however,

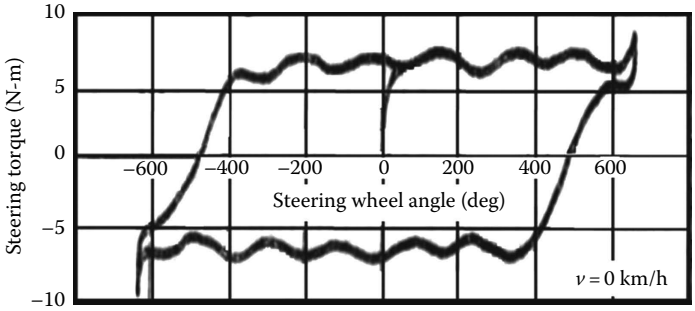
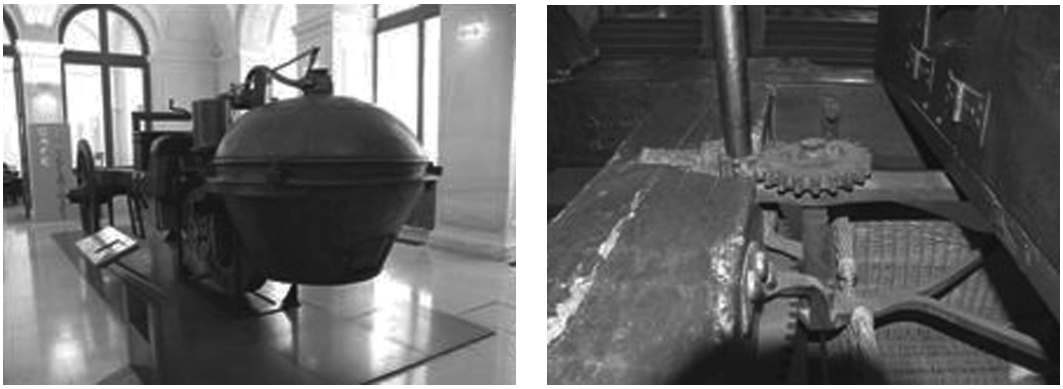


FIGURE 25.2 Relationship between steering torque and steering wheel angle (with power steering system). (From Kageyama, K. et.al., *Vehicle Dynamics*, Riko Press, Tokyo, Japan, 1984.)



(a) (b)

FIGURE 25.3 Gun carriage constructed by Nicolas–Joseph Cugnot. (Courtesy of National Technical Museum, Paris, France.) (a) The gun carriage. (b) Steering system.



FIGURE 25.4 Replica of the automobile drawn by Leonardo da Vinci. (Courtesy of Leonard da Vinci Museum, Florence, Italy.)

four-wheeled vehicles are more effective, and Daimler cars, which were invented the same year, were four wheeled, and adopted the horse-drawn carriage steering mechanism described earlier, which minimizes sideways sliding during geometric turning.

When automobile stability at high speeds is considered, however, these two devices are unsuitable. With horse-drawn-carriage-type steering systems, a moment acts on the side to which the steering is applied. The moment occurs on the steering system shaft rotation and is

due to changes in the lateral force caused by the load transfer between the left and right wheels during turning. Consequently, the link-type steering mechanism, which was invented and patented by Rudolph Ackerman in 1817, came to be adopted. This mechanism performs steering using these links by mounting a rotating shaft to independent left and right wheels (see Chapter 21) and steers the inside wheel much more than the outside one due to the trapezoid links. This mechanism had already been mounted to carriages by Maybach, who invented it in 1889. This mechanism steered the tires using direct links from the steering wheel shaft, but later various steering gears such as the worm gear, planetary gear, screw nut gear, spur gear, and cam lever gear were invented from around 1900 with the objective of reducing the steering force and making steering easier.

Thereafter, the worm roller (cam roller) gear, which reduced friction due to rolling contact by using a single cam roller, was invented, and went on sale commercially in 1926. Almost immediately thereafter, the two-cam roller and three-cam roller were developed and widely adopted.

In 1939, the ball screw gear, with numerous balls inserted between the worm and nut to reduce friction, was invented. Due to problems of machining accuracy, however, this did not become widespread immediately. In 1967, the VGR, which aimed to improve high-speed linearity and reduce steering torque at low speeds by reducing the gear ratio appended to in-line steering wheels and increasing the gear ratio around large corners, was adopted for vehicles, and spread throughout the 1970s.

Further, in 1975, steering systems such as the worm pin type and worm roller type disappeared, and the ball screw type became mainstream in America, Japan, and other countries.

In addition, rack-and-pinion-type steering systems, which have simple link mechanisms, have excellent mountability, and can obtain highly rigid and guarded steering, started to be used in Europe from around 1905 and thereafter became widely adopted from around 1955–1975, such that this format was used overwhelmingly and became mainstream. Due to increasing miniaturization and compaction, and the take-up of four-wheel drive in particular, there was a move from ball screw gears to rack-and-pinion gears. Further, this mechanism, with its rack-and-pinion gear ratio, came to be adopted from around 1981 with the objective of reducing driving torque, leading to the present day.

25.3.2 History of Power Steering

Steering torque increased with the adoption of low-pressure tires, increased weight, and vehicle miniaturization. It was planned to increase gear ratios as a result of this, but power assistance came to be demanded, by heavy vehicles in particular, due to the relationship between the suitability of the steering gear ratio and the steering torque from the aspect of steerability.

In America, in 1926, the hydraulic power steering (PS) was created and was displayed mounted to the Pierce-Arrow Car. Thereafter, various types of power steering, such as the vacuum type, compressed-air type, and electrical type, etc., were tried. From around 1940, hydraulic power steering was mounted to heavy trucks and buses, becoming actualized. Further, due to power steering miniaturization, integral power steering (IPS), which combined the power cylinder and control valve on the steering gearwheel as one unit, was developed in 1962, and went into mass production in 1965.

Meanwhile, hydraulic power steering started to be mounted to ball screw steering systems in passenger vehicles in America from 1951. This ball screw IPS spread, and in America, in 1965, approximately 60% of vehicles had power steering mounted. In Europe, on the other hand, the rack-and-pinion type spread, and in 1955, this type of power steering went into mass production.

After the start of the 1980s, together with the increase in front-wheel load four-wheel drive, the mounting of power steering to ordinary people carriers increased rapidly. Concomitant with the spread of four-wheel drive vehicles, there was a shift from the existing ball screw IPS to rack-and-pinion power steering, and thereafter this type of power steering became mainstream.

In vehicles with a fixed amount of power assistance, there is a tendency for the steering torque to be reduced too much when driving at high speeds. Consequently, steering torque variable power steering soon became necessary in consideration of the balance between steerability and stability.

Around 1966, the amount of output from oil pumps was reduced in response to the number of engine rotations using flow control valves, and variable steering torque power steering that was made sensitive to the number of engine rotations and could suppress the amount of power assistance during high-speed driving. In Europe, in 1970, a system that varied the steering torque according to the vehicle speed and steering angle was mounted to Citroen SM vehicles.

Along with the progress made in electronics technology, during the 1980s, several types of electronically controlled variable torque power steering were developed and adopted one after the other. In 1981, the controllable flow-type electronically controlled variable power steering was adopted, and the controllable reaction force-type electronically controlled variable power steering was adopted in 1983, and then in 1989, the valve-sensitive adjustable control type was also adopted.

In 1988, power steering became electrically powered (EPS). A DC motor, as its power source, was successfully applied to compact vehicles. Compared to the hydraulic type, this EPS, with its simple and lightweight configuration, was an effective system with extremely small engine power loss. Thus, it gradually started to be used in big vehicles as well.

25.4 BASIC THEORY OF STEERING SYSTEMS

25.4.1 Ackerman Steering

The ratio between the steering wheel angle and the front tire steering angle (the mean between the steering angle of the inner wheel and that of the outer wheel) is called the overall steering gear ratio. The size of this value is related to the steerability and stability. As described earlier, and as shown in Figure 25.5, the actual steering angle between the inner and outer wheels is different as shown in Figure 25.6 [3].

The basic ideas for this occurrence are as follows:

With turning at extremely low driving speeds, the centripetal acceleration is almost zero, so there is no need for force on any of the tires. With turning at nonvanishing speed, the forces at the tires are nonvanishing due to the centrifugal force. If such force is applied, slip will occur on the tires. To prevent such tire force being applied improperly to the tires, that is, to prevent friction wear, a common method to set the tires is depicted in Figure 25.5. It should be noticed

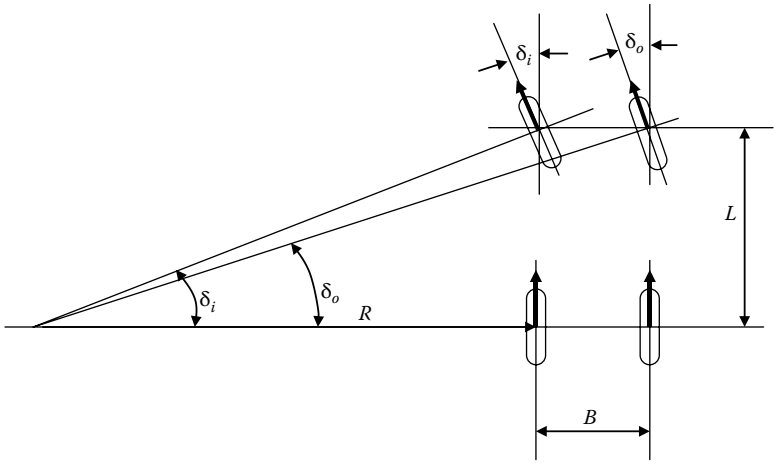


FIGURE 25.5 Geometrical relation between turning radius and steer angle. (From Kondo, M., *Basic Automotive Engineering*, Vol. I, Yokenndo, Tokyo, Japan, 1965; Kageyama, K. et.al., *Vehicle Dynamics*, Riko Press, Tokyo, Japan, 1984; Genta, G., *Motor Vehicle Dynamics*, World Scientific, Singapore, 1997.)

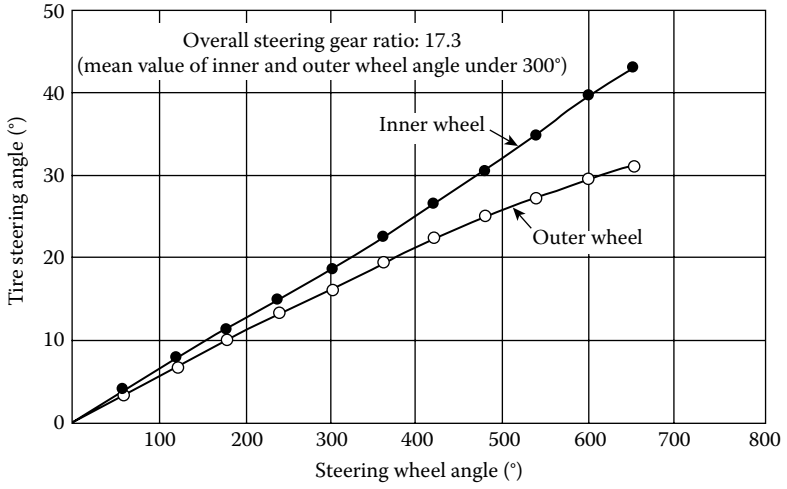


FIGURE 25.6 Example of the relationship between steering wheel angle and tire steering angle. (From Kageyama, K. et.al., *Vehicle Dynamics*, Riko Press, Tokyo, Japan, 1984.)

that this setting (Ackermann) is correct only when the speed is vanishing. When the speed is not vanishing, the setting in Figure 25.5 usually implies almost the same lateral slip at the two wheels of the same axle, which implies that, due to load transfer, the lateral forces are not equal between the two wheels of the same axle (the lateral force of the outer wheel is higher). This occurrence can be overcome by setting different steering angles at each wheel. It is necessary to set the tires so that side slip angle occurs. The position of the rotation center for the vehicle is located on the right-angled direction to each wheel center velocity. The geometrical rotation center at zero speed is on a line extending from each wheel axle. As shown in Figure 25.5, the inside steering angle is called δi, and the outside steering angle is called δo. It is assumed here that the wheel base (L) is the same for both left and right and that the tread (B) is also the same for both front and rear axles. The distance from the rotation center to the inside of the rear wheels is taken to be the turning radius (R). These relationships can be collated and expressed using the following equation:

$$\cot \delta_o - \cot \delta_i = \frac{B}{L} \tag{25.1}$$

If the turning radius here is extremely large compared to the tread, gathering the right and left wheels in the chassis longitudinal axis and assuming the steering angle to be δ, δ can be expressed as described in the following:

$$\tan \delta = \frac{L}{R} \tag{25.2}$$

Next, the relationship between the coordinates fixed on the road surface can be expressed as described in Figure 25.6. The coordinates of the vehicle longitudinal axis fixed on the road surface can be expressed here using the following equation:

$$R = \frac{-\sqrt{\left\{1 + (dy/dx)^2\right\}^3}}{d^2y/dx^2} \tag{25.3}$$

25.4.2 Modeling of the Steering System

25.4.2.1 Equivalent Moment of Inertia in Steering Systems

Steering systems perform steering using steering gears, so it is necessary to examine the equivalent moment of inertia around the steering shaft. A model of the steering system is described in Figure 25.7. Assuming that screw displacement is small, the equation of equilibrium in this diagram can be determined using the following equations [3]:

$$\begin{aligned} T &= I_h\ddot{\theta}_h + r_1F_1 \\ r_2F_1 &= I_s\ddot{\theta}_s + r_3F_2 \\ F_2 &= m\ddot{x} \end{aligned} \tag{25.4}$$

Further, the following formula describes the relationship between the angles and the displacement:

$$\begin{aligned} r_1\theta_h &= r_2\theta_s \\ r_3\theta_s &= x \end{aligned} \tag{25.5}$$

Consequently, the following formula can be obtained from these relationships:

$$T = \left\{ I_h + \left(\frac{r_1}{r_2}\right)^2 \left(I_s + r_3{}^2 m\right) \right\} \ddot{\theta}_h \tag{25.6}$$

Consequently, the equivalent moment of inertia around the steering shaft of the steering system can be obtained using the following formula:

$$I_{eh} = I_h + \left(\frac{r_1}{r_2}\right)^2 \left(I_s + r_3{}^2 m\right) = I_h + \frac{\left(I_s + r_3{}^2 m\right)}{n^2} \tag{25.7}$$

Here, n describes the gear ratio ($n = r_2/r_1$).

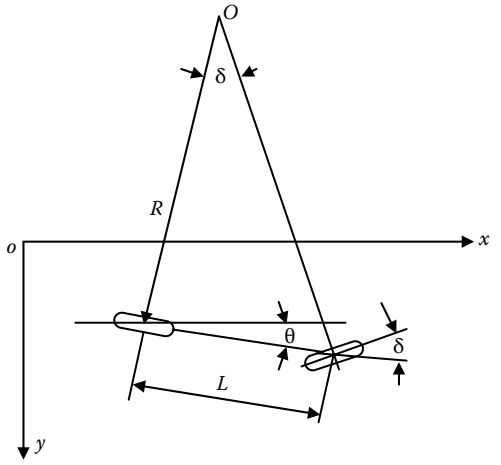


FIGURE 25.7 Geometrical relation between turning radius and steer angle. (From Kageyama, K. et.al., *Vehicle Dynamics*, Riko Press, Tokyo, Japan, 1984; Genta, G., *Motor Vehicle Dynamics*, World Scientific, Singapore, 1997.)

Conversely, the moment around the shaft behind the steering gear can be described similarly as shown by the following equation:

$$I_{es} = I_h n^2 + I_s + r_3^2 m \tag{25.8}$$

From the preceding equation, it can be seen that if passing through the steering gear (gear ratio n), the moment of inertia for the steering wheel are $x\,n^2$ over the circumference of the front-wheel steering shaft.

Simplifying this analysis means considering the rotation operation around the kingpin only. The equivalent moment of inertia for the steering system described in Figure 25.7 can be determined as described in the following equation:

$$I_{es} = I_h n^2 + I_s \tag{25.9}$$

25.4.2.2 Modeling for Simple Steering Systems

The steering system needs to be modeled by considering viscous friction and dry friction. The friction occurs in the shaft, and the elastic warping of the shaft is also important. The steering performed by the driver on the vehicle becomes angle inputs or torque inputs. The actual steering angle of the front tires here is not proportionally related to the steering wheel angle. To consider these effects, it is necessary to configure the equations of motion that describe the steering system, and to solve them simultaneously with the vehicle equations of motion. This section instructs on the equations of motion for the steering system only.

The steering system in Figure 25.8 has two degrees of freedom, and the equations of motion for the steering torque inputs can be solved simultaneously with the equation of equilibrium of the moments for the kingpin circumference and the balance for the moments of the steering shaft circumference. For simplification, try replacing the item on the steering shaft circumference with the kingpin shaft circumference.

Balance of steering shaft circumference:

$$I_h'\ddot{\theta}_h' + D_s\theta_h' + k\theta_h' - k\theta_s - D_s\theta_s = T_h' \tag{25.10}$$

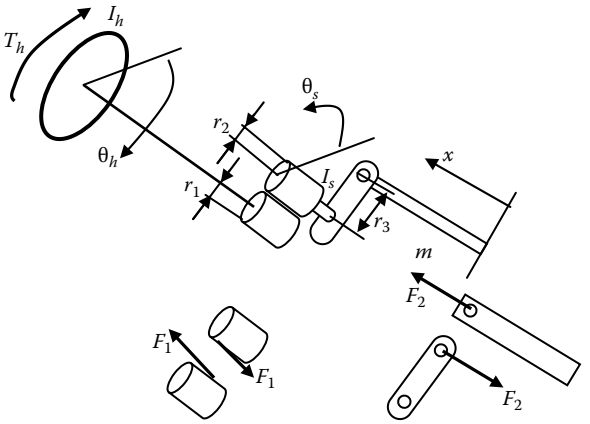


FIGURE 25.8 Element of steering system.

Balance of kingpin shaft circumference:

$$I_s\ddot{\theta}_s + D_s\dot{\theta}_s + k\theta_s - k\theta'_h - D_s\dot{\theta}'_h = -T_s \tag{25.11}$$

With these formulae, the parameters marked with an apostrophe (′) describe the values calculated for the kingpin shaft circumference and are expressed using the following formula:

$$I'_h = I_h n^2, \quad T'_h = T_h n, \quad \theta'_h = \frac{\theta_h}{n} \tag{25.12}$$

Further, the parameters are also described in Table 25.1.

With steering wheel angle inputs, it is not necessary to consider the effect of moment of inertia for the steering wheel. Therefore, the simple equation for the steering system is described as one degree of freedom around the kingpin shaft, and the total equations of motion are solved with the vehicle equations of motion simultaneously:

$$k\theta'_h - T_s = I_s\ddot{\theta}_s + D_s\dot{\theta}_s + k\theta_s \tag{25.13}$$

In this equation, T_S is the restoration moment of the tire and is the product of adding the self-aligning torque and the moment caused by front-wheel geometry (i.e., the effects of caster trail, kingpin inclination, roll center, kickback from road, and so on) (Table 25.2).

TABLE 25.1

Parameter	Explanation
T_h	Steering torque
I_h	Moment of inertia around steering axis
I_s	Moment of inertia around output steering gear
θ_h	Steering wheel angle
θ_s	Steer angle for output gear
F_1	Inner force of steering gear
F_2	Inner force of steering link
m	Mass for link
r_1	Radius for input steering gear
r_2	Radius for output steering gear
r_3	Radius for pitman arm

TABLE 25.2

Parameter	Explanation
T'_h	Equivalent steering torque
I'_h	Equivalent moment of inertia around
I_s	Moment of inertia around kingpin
θ'_h	Equivalent steering wheel angle around kingpin
Δ	Steer angle
K	Rotational spring constant around steering shaft
T_s	Restoration moment
D_s	Viscous resistance coefficient
K_s	Cornering stiffness
B	Slip angle
V	Vehicle speed

25.4.2.3 Detailed Model of the Steering System

The preceding section examined a simple model of the steering system. Here, the steering system described is a more advanced model [4].

As described in Figure 25.9, the steering shaft has an inclination of Φ over the vertical shaft. Consequently, the vehicle rotation operation is affected. Here, the yawing velocity must be considered. Further, the moments and torque due to viscous resistance and dry resistance in the bearings are used. In addition, it is also necessary to consider the elastic warping of the parts that maintain the steering system. Moreover, when the front wheels rotate, and the camber angle changes in the steering wheels due to the roll operation, the gyro moment affects the steering shaft circumference. This section describes the operational formulas for steering systems that consider these points.

In Figures 25.9 and 25.10, the steering system can be divided into two parts, above and below the steering gear. The operational formula from the steering wheel to the steering gear can be described using the following equation:

$$T_h = I_h(\ddot{\theta}_h + \dot{r}\cos\phi) + k_h\left(1 - \frac{n^2 k_h}{n^2 k_h + k_s + k_1}\right)\theta_h - \frac{nk_h k_s}{n^2 k_h + k_s + k_1}\theta_s \pm \frac{n^2 k_h}{n^2 k_h + k_s k_1} R_h \tag{25.14}$$

The parameters used here are described in Table 25.1. Further, in this formula, the viscous resistance of the shaft is not considered. Next, the operational formula from the steering gear to the front wheels can be described using the following equation:

$$I_s(\ddot{\theta}_s + \dot{r}) - \frac{nk_h k_s}{n^2 k_h + k_s + k_1}\theta_h - \left(\frac{{k_s}^2}{n^2 k_h + k_s + k_1} - k_s - k_2 - 2K_s\right)\theta_s + D_s\dot{\theta}_s + 2K_s\beta$$
$$+ 2I_w\frac{v}{R}\dot{\gamma} \pm \left(\frac{n^2 k_h}{n^2 k_h + k_s k_1} R_h - R_s\right) = 0 \tag{25.15}$$

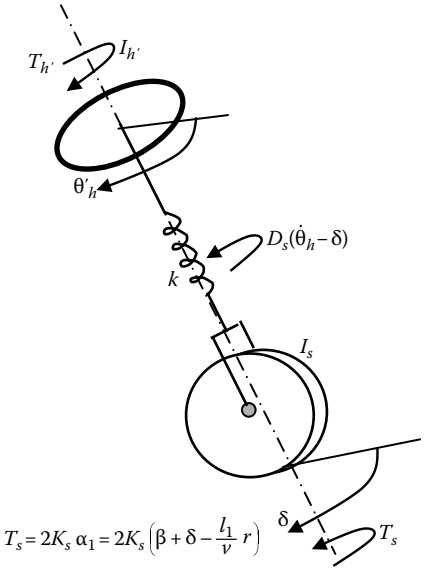


FIGURE 25.9 Simple steering model.

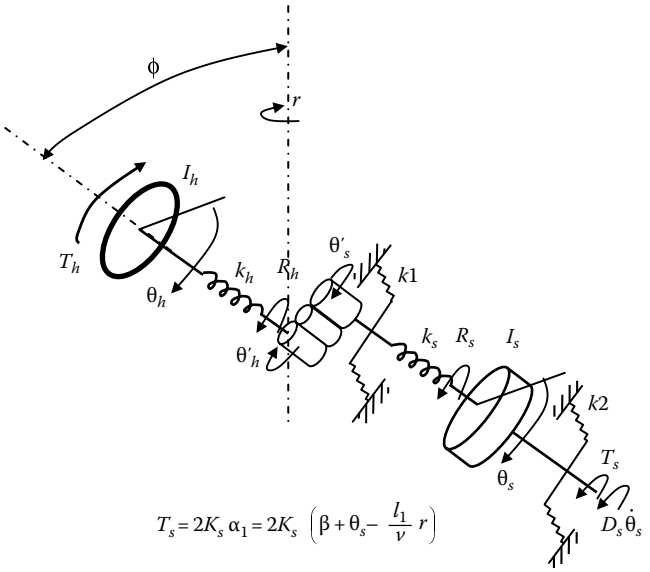


FIGURE 25.10 Detailed model of the steering system. (From Kondo, M., *Basic Automotive Engineering*, Vol. II, Yokenndo, Tokyo, Japan, 1967.)

TABLE 25.3

Parameter	Explanation
Φ	Inclination of steering shaft
R	Yawing velocity
K_h	Rotational spring constant around steering shaft
K_s	Rotational spring constant around kingpin
k_1	Spring constant at installation part 1
k_2	Spring constant at installation part 2
D_s	Viscous resistance coefficient around kingpin
R_h	Dry friction around steering shaft
R_s	Dry friction around kingpin

In this formula, the gyro moments due to the rotation of the steering wheels are considered. The section on viscous resistance considers the kingpin shaft circumference only. Further, the ± in the formulas is the section on dry friction, and tends to reverse rotation according to the direction of operation and rotation of each part (Table 25.3).

25.5 POWER STEERING

To lighten the steering torque in manual steering, it is necessary to increase the steering gear ratio and reduce the friction in all parts. By increasing the gear ratio, however, the steering angle is also increased at the same time. Power steering devices are steering systems that are equipped with motive power assistance devices to assist the steering torque, with the aim of reducing the steering torque.

By using this power assistance, the gear ratio can be selected freely, and the amount of design freedom can also be increased. Steering is a bilateral system, and as it is not the driver who inputs the steering torque but that the torque is input to the steering system from the tires, for the driver,

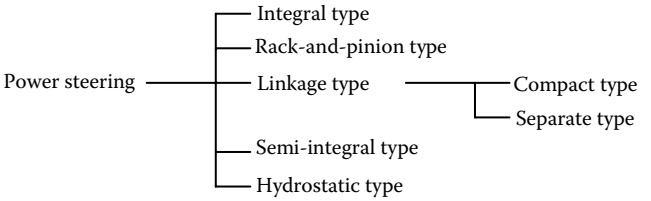
this also becomes a type of information while driving. With this in mind, it is necessary to consider power assistance that satisfies the following requirements [5]:

- Steering torque is suitable.
- Both steering torque and maintained steering force enable the close relationship with the road surface reaction force to be felt.
- There is no impediment to recovery from turning the vehicle's steering wheel.
- This enables suitable absorption of shock from the road surface.

25.5.1 Types of Power Steering

Hydraulics, air pressure, and electric motor are all sources of motive power in the power assistance devices of power steering, but the hydraulic type is the mainstream power steering due to its compactness, high output, responsiveness, and reliability. Further, in recent years, due to its excellent ease of mounting, attributed to its lightweight compactness and simple configuration, the electric type power steering has started to be adopted, mainly for compact vehicles [5].

There are various types of hydraulic power steering, depending on the control valve; power cylinder; and steering gear box configuration, arrangement, and assembly. These can be classified as follows:

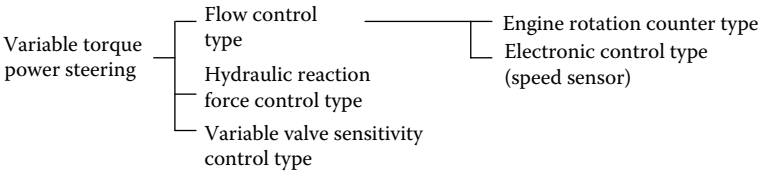


Classifications are shown in the preceding diagram.

The integral type combines the control valve and power cylinder inside a ball-screw-type steering gear box and is used in some passenger vehicles, trucks, and buses. The rack-and-pinion type combines the control valve and power cylinder inside this type of gear box, and the main unit is combined with part of the link mechanism. This type is widely used in passenger vehicles. The linkage type arranges the control valve and power cylinder in the link part, which is behind the steering gears. This is either all-in-one or separate, and can be classified either as combined or separate type. This type of power steering is historically old and is currently used only in some large trucks. The semi-integral type combines the control valves in the gear box, and the power cylinder is mounted to the link part. This type is used in industrial vehicles and construction vehicles. The hydrostatic type operates the power cylinders for the car wheels using hydraulics from the control valve, which is mounted to the operation part, with no mechanical link connections between the operation part and the wheels. This type is used in agricultural tractors, fork lift trucks, and construction vehicles.

With power steering, it is necessary to control the amount of power assistance according to the conditions. Essentially, steering torque becomes excessive and power assistance is necessary when accelerating at extremely low speeds from a standstill. Further, conversely, if excessive assistance is performed during high-speed driving, the vehicle operations will become oversensitive, and there is a tendency for the steering wheel system stability to be lost. Here, it is necessary to perform major power assistance at extremely low speeds in garages and during major steering maneuvers and to perform controls to reduce assistance at high speeds. Normally, control by

speed sensors and by engine rotation counters is performed. These can be collated as shown in the following diagram.



25.5.2 Modeling for Power Steering Systems

As shown in the aforementioned diagram, with hydraulic power steering, which is the current mainstream and which performs power assistance using controlled operating oil, the power cylinders can be divided into three parts: control valves, which control the direction of the hydraulic flow, and oil pumps, which are powered by the engine. A torsion bar is mounted to the steering shaft, and turn torque is created in the steering shaft by steering, and the direction of the operating oil is controlled by operating the control valve (e.g., spool valve) according to the turn angle of this torsion bar. This operating oil that is being controlled is fed into the power cylinder to enable power assistance to be performed.

The parameters are described in Table 25.4.

The formula for balancing the circumference of the steering wheel is described as follows:

$$I_h\ddot{\theta}_h + C_h\dot{\theta}_h + K_t(\theta_h - \theta_w) = T_h \tag{25.16}$$

TABLE 25.4

Parameter	Explanation
T_h	Steer torque around steering wheel
T_w	Warm shaft torque
T_c	Sector torque
F_p	Force between warm shaft and ball-nut
I_h	Moment of inertia around steering wheel
I_w	Moment of inertia around warm shaft
I_c	Moment of inertia around sector shaft
M_p	Mass for ball-nut
Θ_h	Rotational angle around steering shaft
Θ_w	Rotational angle around warm shaft
Θ_c	Rotational angle around sector shaft
X_p	Lateral deviation of ball-nut
K_t	Equivalent rotational spring constant for spool
K_c	Equivalent rotational spring constant for output shaft
C_h	Viscous resistance coefficient around steering shaft
C_w	Viscous resistance coefficient around warm shaft
C_p	Viscous resistance coefficient of ball-nut
C_c	Viscous resistance coefficient around output shaft
R_c	Radius of pitch circle of sector gear
E_w	Lead of warm gear
A_p	Effective area for ball-nut
P_a	Pressure of ball-nut vessel *a*
P_b	Pressure of ball-nut vessel *b*

The formula for balancing the sector circumference is as follows:

$$I_w\ddot{\theta}_w + C_w\dot{\theta}_w + K_t(\theta_w - \theta_h) = -\left(\frac{e_w}{2\pi}\right)F_p \tag{25.17}$$

The formula for balancing the ball-nut circumference is as follows:

$$m_p\ddot{x}_p + C_p\dot{x}_p = F_p + A_p(p_a - p_b) - \frac{T_c}{r_c} \tag{25.18}$$

The following formula can be obtained using the balance with the external force around the output shaft (sector shaft):

$$I_c\ddot{\theta}_c + C_p\dot{\theta}_c - K_c\theta_c = T_c \tag{25.19}$$

Here

$$\begin{aligned} x_p &= \frac{e_w}{2\pi}\theta_w \\ \theta_c &= \frac{x_p}{r_c} = \frac{\theta_w}{n} \\ n &= \frac{2\pi r_c}{e_w} \end{aligned} \tag{25.20}$$

From the preceding relationships, the equations of motion for the power steering system can be written as follows:

$$I_h\ddot{\theta}_h + C_h\dot{\theta}_h + K_t(\theta_h - \theta_w) = T_h \tag{25.21}$$

$$I_{wt}\ddot{\theta}_w + C_{wt}\dot{\theta}_w + K_{ct}\theta_w + K_t(\theta_w - \theta_h) = A_{pt}(p_a - p_b) \tag{25.22}$$

Here

$$\begin{aligned} I_{wt} &= I_w + \left(\frac{r_c}{n}\right)^2 m_p + \left(\frac{1}{n}\right)^2 I_c \\ C_{wt} &= C_w + \left(\frac{r_c}{n}\right)^2 C_p + \left(\frac{1}{n}\right)^2 C_c \end{aligned} \tag{25.23}$$

$$K_{ct} = \left(\frac{1}{n}\right)^2 K_c$$

$$A_{pt} = \frac{r_c}{n}A_p$$

25.5.3 Example Analysis Including Power Steering System

An analysis using the power steering system model described in the preceding paragraph is shown. The block diagram of the power steering system is shown in Figure 25.11.

Figure 25.12 shows an analytical result of frequency response of the vehicle using the model in Figure 25.11. The parameter of this calculation is rigidity of the power steering system, and the vehicle used in the calculation is a light truck.

From this figure, it is found that the rigidity of the PS has a big influence on the vehicle behavior [5]. And, the influence of sensitivity of the valve and the viscous resistance inside the PS to vehicle behavior has been examined [5–7].

In addition, it is important that the evaluation method of the PS has been investigated from the viewpoint of human engineering [8–12]

25.5.4 Electric Power Steering

The electric power steering (EPS) system has high potential because various characteristics can be reproduced by using its control algorithm [11,13–17]. Furthermore, various driver support systems are constructed by using the EPS, for example, parking assistance, lane keeping, and so on [18–21].

Moreover, this system can attempt the improvement of fuel consumption compared with the hydraulic-type power steering system. Therefore, many compact cars have mainly adopted the EPS

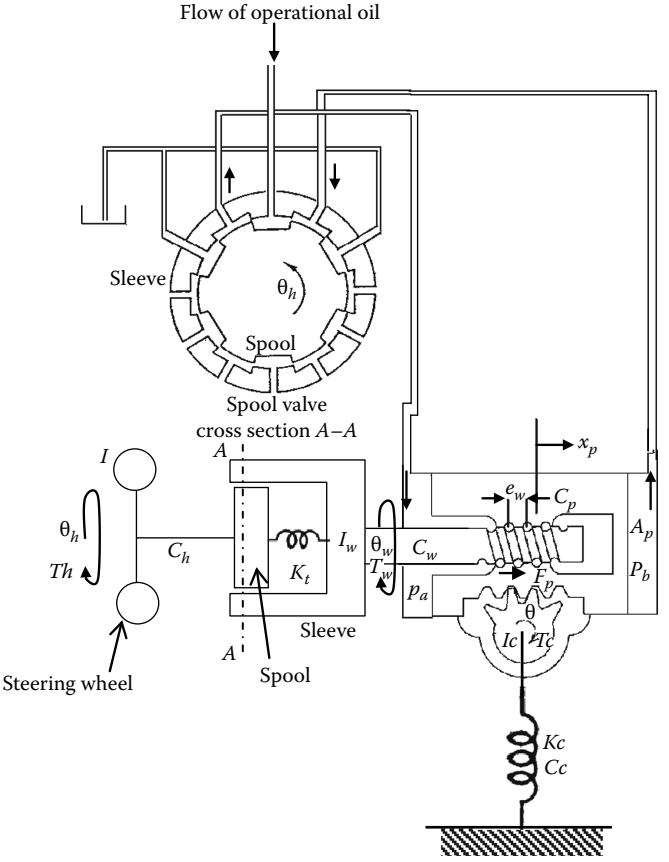


FIGURE 25.11 Modeling of power steering system. (From Komamura, S. et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.)

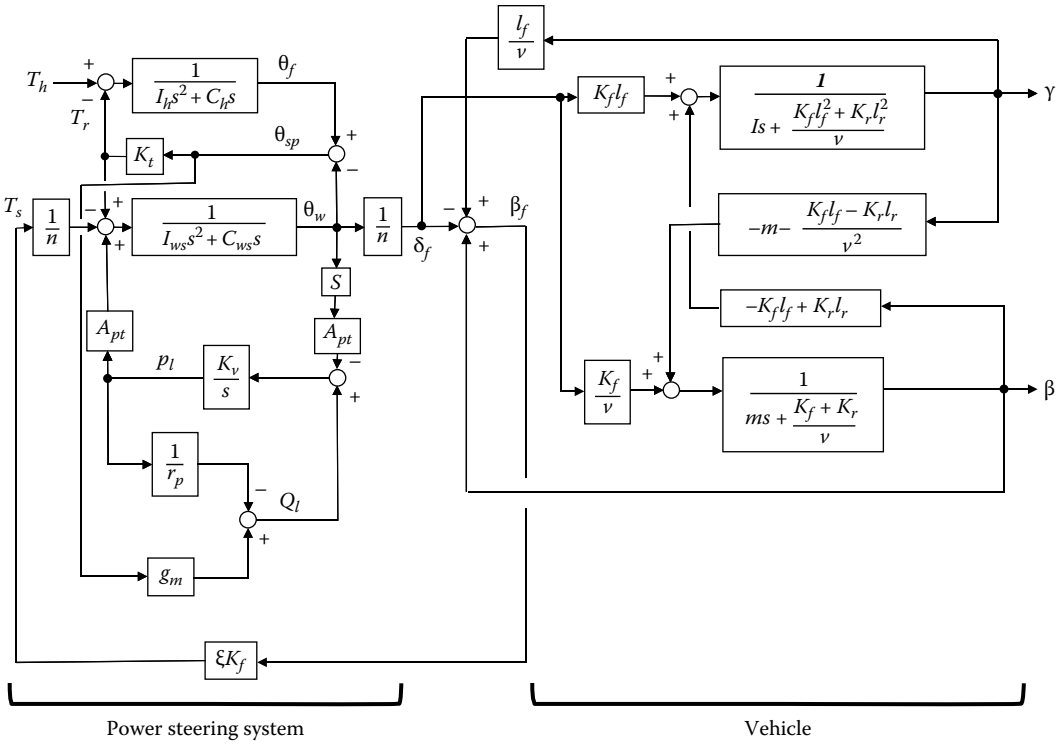


FIGURE 25.12 Block diagram for power steering system. (From Komamura, S. et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.)

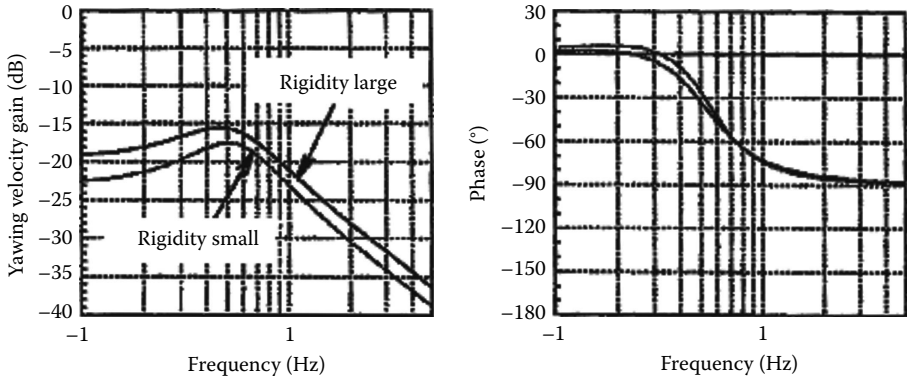


FIGURE 25.13 Analytical result of the vehicle including power steering (Influence of rigidity of PS). (From Komamura, S. et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.)

in recent years [5]. The EPS is divided roughly into the following three types by the assist position of a motor: the column-assisted type, the pinion-assisted type, and the rack-assisted type as shown in Figure 25.13.

25.5.4.1 Modeling for EPS

Recently, many EPSs in the market adopt the pinion-assist type. Therefore, the pinion-assist-type EPS is described as the analytical model (see Figure 25.14) [16].

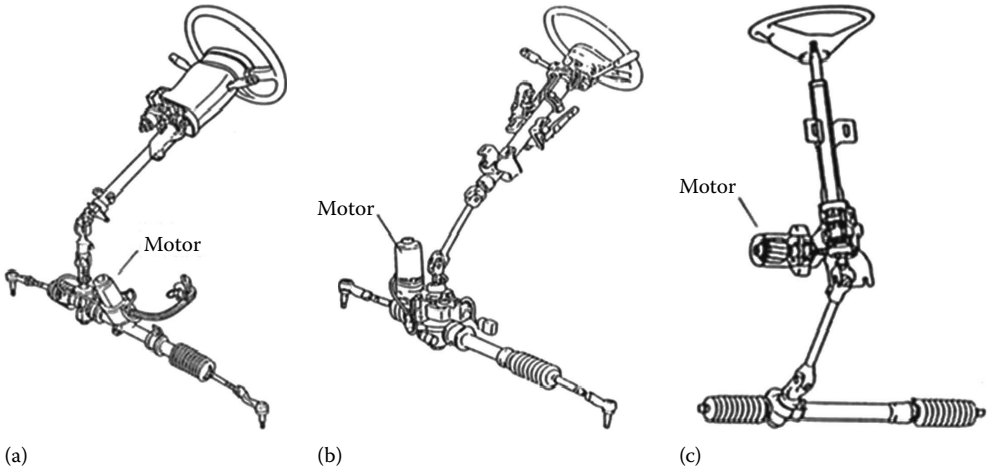


FIGURE 25.14 EPS types. (a) Rack-assist type, (b) Pinion-assist type, (c) Column-assist type. (From Komamura, S. et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.)

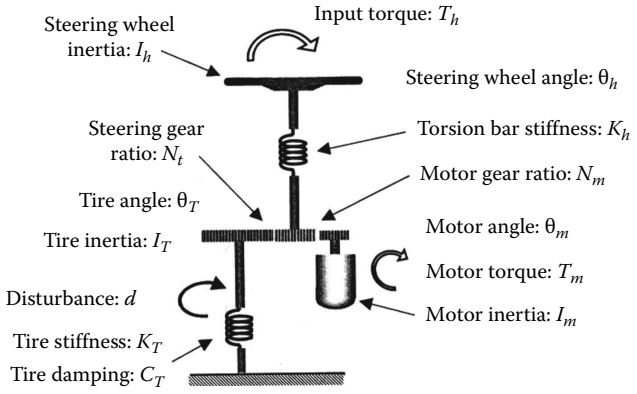


FIGURE 25.15 Model of electric power steering. (From Takehara, S., Control of electric power steering, Preprint of intensive seminar for vehicle dynamics, *JSME*, 08–44, 45, 2008.)

The parameters used here are described in Figure 25.14. Further, in this model, the viscous resistance of the model is considered in the part of tire. Next, this model is simplified as shown in Figure 25.15. In this model, the equivalent parameters are used as shown in Figure 25.15.

Using this model, the equations of motion for the EPS are described as follows:

$$I_H\ddot{\theta}_H + K_H(\theta_H - \theta_M) = T_H$$
$$I_M\ddot{\theta}_M + C_T\dot{\theta}_M + K_T\theta_M + K_H(\theta_M - \theta_H) = T_M + d \tag{25.24}$$

25.5.4.2 Block Diagram of Control System

An analysis using the electric power steering model described in the preceding paragraph is shown. Therefore, the block diagram of the control system for the EPS is as shown in Figure 25.16. In this figure, K_1 and K_2 are the control gains: K_1 is the assist gain and K_2 is the damping gain for the total system [16].

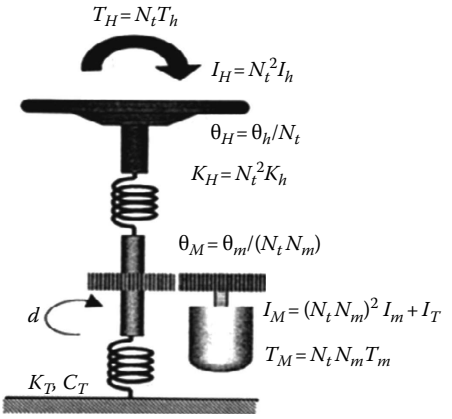


FIGURE 25.16 Simplified electric power steering model. (From Takehara, S., Control of electric power steering, Preprint of intensive seminar for vehicle dynamics, *JSME*, 08–44, 45, 2008.)

25.5.4.3 Vehicle Model Including EPS

An analytical vehicle model with the simplified EPS is described as shown in Figure 25.17. Here, the caster trail, which is the length between kingpin axis at the road and the center of a tire contact point, is described by ξ as shown in Figure 25.1. The parameters of the vehicle are shown in Figures 25.17 and 25.18 [16].

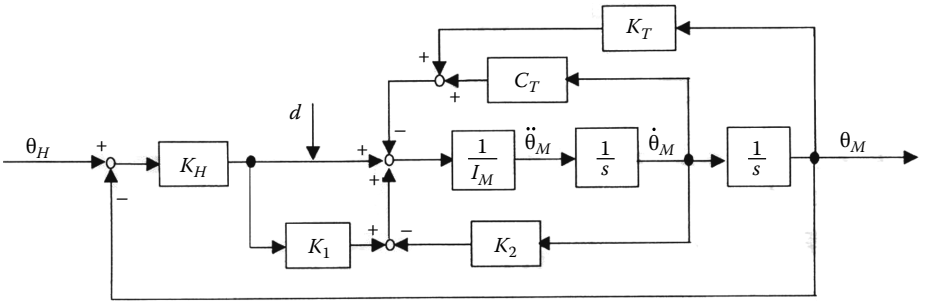


FIGURE 25.17 Block diagram of control system with EPS. (From Takehara, S., Control of electric power steering, Preprint of intensive seminar for vehicle dynamics, *JSME*, 08–44, 45, 2008.)

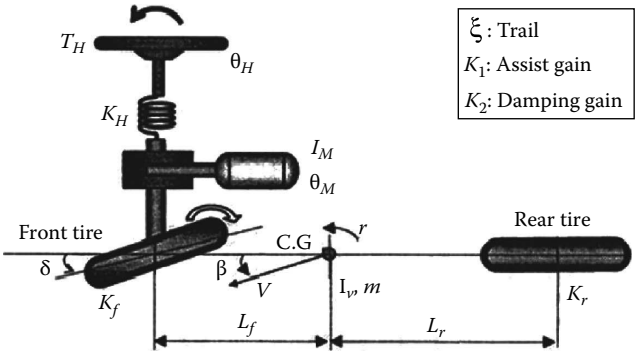


FIGURE 25.18 Vehicle model with EPS. (From Andonian, B. et al., Driver steering performance using joystick vs steering wheel controls, SAE Paper No. 2003-01-0118, 2003.)

The formulas for balancing the circumference of the vehicle are described as follows:

$$\left\{mv + \frac{2(l_f K_f - l_r K_r)}{v}\right\} r + mv\dot{\beta} + 2(K_f + K_r)\beta = 2K_f\delta$$

$$I_v\dot{r} + \frac{2\left(l_f{}^2 K_f + l_r{}^2 K_r\right)}{v} r + 2(l_f K_f - l_r K_r)\beta = 2l_f K_f\delta \tag{25.25}$$

The equations of the EPS part are described as follows:

$$I_H\ddot{\theta}_H + K_H(\delta - \theta_H) = T_H$$

$$I_M\ddot{\delta} + C_T\dot{\delta} + 2\xi K_f\delta + K_H(\delta - \theta_H) - 2\xi K_f\left(\beta + \frac{l_f r}{v}\right) = T_M \tag{25.26}$$

$$T_M = K_1 K_H(\theta_H - \theta_M) - K_2\dot{\theta}_M$$

Using these equations, the relationship between vehicle behavior and design parameter of the EPS is derived [5,21,16]. In addition, it is important that the evaluation method of the EPS is investigated from the viewpoint of human engineering, and also the examination of robustness for the control system becomes important [16,17].

REFERENCES

1. NHK Japan (Ed.), *Automobile—What Has the Human Race Made*? No. 2, Japan Broadcasting (NHK), 1980.
2. G. Genta, *Motor Vehicle Dynamics*, World Scientific, Singapore, 1997.
3. M. Kondo, *Basic Automotive Engineering*, Vol. I, Yokenndo, Tokyo, Japan, 1965.
4. M. Kondo, *Basic Automotive Engineering*, Vol. II, Yokenndo, Tokyo, Japan, 1967.
5. S. Komamura et al., *Steering and Dynamics*, Sankai-do Press, Tokyo, Japan, 1996.
6. S. Nishimura et al., Analysis of response delay during rapid steering in hydraulic power steering system, *Proc. JSAE*, No. 62, 1998.
7. G.E. Smid, M.S. Qatu, and J. Drew, Optimizing the power steering components to attenuate noise and vibrations, *European Conference on Vehicle Noise and Vibration*, IMechE, London, U.K., 1998.
8. T. Hara et al., Driver's steering quality in a monotonous driving, *Proceedings of 14th ESV*, Munich, Germany, 1994.
9. A. Isomura et al., Human factors in driver's steering operation, *Trans. JSAE*, 27(1), 1996.
10. Y. Hisaoka et al., A research on desirable steering response and steering torque for driver's feeling, *Trans. JSAE*, 28(4), 1997.
11. J. Tajima et al., Research on effect of steering characteristics on control performance of driver-vehicle system from a viewpoint of steer-by-wire system design, *Proceedings of AVEC'98*, Nagoya, Japan, 1998.
12. M. Segawa et al., Future technical issue on SBW system as the human-machine interface, *J. JSAE*, 57(5), 2003.
13. Y. Shimizu et al., Effects and design of steering system with variable gear-ratio according to vehicle speed and steering wheel angle, *Proc. JSAE*, 21–99, 9–12, 1999.
14. S. Nakano et al., A study of assist control strategy for electric power steering system, *Trans. JSAE*, 31(2), 47–52, 2000.
15. M. Segawa et al., A study on the relationship between vehicle behavior and steering wheel torque on steer by wire vehicles, *Vehicle Syst. Dyn*., 41, 202–211, 2004.
16. S. Takehara, Control of electric power steering, Preprint of intensive seminar for vehicle dynamics, *JSME*, 08–44, 45, 2008.
17. Y. Oniwa. and Y. Shimizu, Evaluation method for steer assist feeling around steering center and control design on EPS, *J. Mech. Syst. Transportation Logistics*, 1(1), 2008.

18. S. Nakano et al., Steering Control Strategies for Steer-by-Wire System (Second Report), *Trans. JSAE*, 33(3), 121–126, 2002.
19. K. Kanai, Drive-by-wire of automobile-symbiosis with airplane, *J. JSAE*, 58(5), 80–87, 2004.
20. A. Heathershaw, Matching of chassis and variable ratio steering characteristics to improve high speed stability, SAE Paper No. 2004–01–1103, 2004.
21. L. Liu, Analysis on electric power steering system to improve vehicle handling and steering feeling for avoidance maneuver, *Proceedings of Control Corroboration Conference*, Vol. 50, 2007.
22. K. Kageyama et.al., *Vehicle Dynamics*, Riko Press, Tokyo, Japan, 1984.
23. B. Andonian et al., Driver steering performance using joystick vs steering wheel controls, SAE Paper No. 2003-01-0118, 2003.
24. A. Suzuki, Analysis and design of servo-system for power steering system, *J. JSAE*, 24(8), 1970.
25. H. Inoue et al., Types and features of steering systems, *J. JSAE*, 31(11), 1977.
26. M. Nishikawa et al., Power-assisted steering with velocity and load-proportional feel, *J. JSAE*, 32(1), 1978.
27. S.Sano et al., Influence of vehicle response parameters on driver control performance, *Proceedings of 18th FISITA*, Hamburg, Germany, 1980.
28. S. Sano et al., Effects of vehicle response characteristics and driver's skill level on task performance and subjective ratings, *Proceedings of ESV*, Washington, DC, 1980.
29. H. Sato et al., Dynamic characteristics of a whole wheel steering vehicle with yaw velocity feedback rear wheel steering, *IMechE*, C124/83, Washington, DC, 1983.
30. T. Kojima, Steering engineering of recent vehicles, *J. JSAE*, 38(4), 1984.
31. Y. Shibahata et al., Effects of rear suspension steer properties on handling and stability, *Proc. JSAE*, No. 815, 1985.
32. S. Sano et al., Four wheel steering system with rear wheel steer angle controlled as function of steering wheel angle, SAE Paper No. 860625, 1986.
33. T. Takiguchi et al., Improvement of vehicle dynamics by vehicle-speed sensing four-wheel steering system, SAE Paper No. 860624, 1986.
34. N. Irie et al., HICAS-improvement of vehicle stability and controllability by rear suspension steering characteristics, *Proceedings of FISITA*, Belgrade, Serbia, 1986.
35. Y. Shibahata et al., The development for an experimental four-wheel-steering vehicle, SAE Paper No. 860623, 1986.
36. Y. Matsushita et al., Modeling of vehicle and improvement of handling quality by adaptive front-wheel steering control system, *Trans. JSAE*, 36, 1986.
37. H. Hayashi, Optimum control of a driver/four-wheel-steered-vehicle system, *Proceedings of 10th IAVSD*, Prague, Czechoslovakia, 1987.
38. K.H. Senger et al., The influence of a four wheel steering system on the stability behavior of a vehicle-driver system, *Proc. 10th IAVSD*, Prague, Czechoslovakia, 1987.
39. P.M. Leucht, Active four-wheel-steering design for the blazer XT-1, SAE Paper No. 872285, 1987.
40. K. Ito et al., A new way of controlling a four wheel steering vehicle, *Trans. SICE*, 23(3), 48–54, 1987.
41. M. Iguchi, A new design concept of vehicle dynamics based on active control, *JSME Int. J*., Series. III, 31, 1–7, 1988.
42. M. Nagai, Control of 4WS systems-now and future, *J. JARI*, 10(5), 1988.
43. M. Nagai et al., Theoretical study on active four-wheel-steering system by virtual-vehicle model following control, *JSAE Rev*., 9(4), 01414847, 1988.
44. F.Sugasawa et al., Improvements in vehicle control and stability through front- and rear-wheel steering control, *Trans. JSAE*, 38, 1988.
45. J.C. Whitehead, Four wheel steering: Maneuverability and high speed stabilization, SAE Paper No. 880642, 1988.
46. M.J. Vanderploeg et al., Evaluation of four-wheel steer path following performance using a linear inverse vehicle model, SAE Paper No. 880644, 1988.
47. K. Fukui et al., Analysis of driver and a four wheel steering vehicle system using a driving simulator, SAE Paper No. 880641, 1988.
48. H. Mouri et al., Handling and stability improvement achieved with four-wheel steering, *IMechE*, C441/88, London, U.K., 1988.
49. A.G. Nalecz et al., Investigation into the stability of four wheel steering vehicles, *Int. J. Vehicle Des*., 9(2), 159–178, 1988.
50. A.G. Nalecz et al., Analysis of the dynamic response of four wheel steering vehicles at high speed, *Int. J. Vehicle Des*., 9(2), 179–202, 1988.

51. J.E. Bernard et al., Linear analysis of a vehicle with four-wheel steering, SAE Paper No 880643, 1988.
52. K.H. Senger et al., A velocity dependent steering gear ratio for four-wheel steered vehicle, *IMechE*, C438/88, London, U.K., 1988.
53. P. Lugner et al., Theoretical investigations on the behavior of a car with additional four-wheel steering at μ-split conditions, *IMechE*, C440/88, London, U.K., 1988.
54. K. Kanai et al., Design of adaptive-type yaw rate, lateral acceleration and D^*-controllers for 4-wheel steering car, *Trans. SICE*, 24(4), 1988.
55. M. Yamamoto et al., Controllability and stability aspects of actively controlled 4WS vehicle, *Proceedings of 11th IAVSD*, Kingston, Ontario, Canada, 1989.
56. H. Takeda et al., A review of four-wheel steering studies from the viewpoint of vehicle dynamics and control, *Vehicle System Dyn.*, 18, 151–186, 1989.
57. T. Fujishiro et al., A study on the stability of a four wheel steering vehicle as a steering torque input system, *Trans. SICE*, 25(2), 1989.
58. S. Sano et al., Effect of control principle of steer angle dependent 4WS on steering response characteristics, *Trans. JSAE*, 43, 1990.
59. M. Yamamoto et al., Improvement of steering response and disturbance response by active controlled rear wheel steer, *Trans. JSAE*, 46, 1990.
60. JSAE, *Automotive Handbook*, JSAE, Tokyo, Japan, 1991.
61. JAMCA, *Structure of Chassis*, JAMCA, 1991.
62. M. Yamamoto, Active control strategy for improved handling and stability, SAE Paper No. 911902, 1991.
63. M. Aga et al., Design of active control system for front-and-rear-wheel steering vehicle, *Trans. JSAE*, 22(2), 1991.
64. Matsunaga et al., Analysis of self-excited vibration in power steering system, *Proceedings of AVEC'92*, Yokohama, Japan, 1992.
65. F. Jian-Guo et al., System modeling and simulation method of electric motor-driven power steering, *Proceedings of AVEC'92*, Yokohama, Japan, 1992.
66. M. Abe, *Vehicle Dynamics and Control*, Sankai-do Press, Tokyo, Japan, 1992.
67. M. Nagai et al., Stability of 4WS systems based on side slip zeroing control, *Trans. JSAE*, 23(1), 1992.
68. E. Adachi et al., Static steering noise and bending vibrations of an automotive power steering system, *Trans. JSAE*, 24(1), 1993.
69. A. Higuchi et al., Optimal control of four wheel steering vehicle, *Vehicle Syst. Dyn.*, 22, 397–410, 1993.
70. T. Shiotsuka et al., Control of a four-wheel steering car by neural network, *Trans. JSAE*, 25(4), 1994.
71. S. Takeuchi and E. Adachi., Analysis and reduction of power steering noise during static steering operation, SAE Paper No. 950581, 1995.
72. N. Yuhara et al., An advanced steering system with active kinesthetic feedback for handling qualities improvement, *Vehicle Syst. Dyn.*, 27, 327–355, 1997.
73. G.R. Ferries and R.L. Arbanas., Control/structure interaction in hydraulic power steering systems, *Proceedings of the American Control Conference*, Albuquerque, NM, 1997.
74. M. Abe et al., Technology for improvement of vehicle handling, *Trans. JSAE*, 1998.
75. I. Yamazaki et al., A study of driving with control stick, *Trans. JSAE*, 29(3), 1998.
76. S. Taheri, R. Kazemi, and M. Tabatabai, Analysis and optimization of vehicle steering system, SAE Paper No. 981113, 1998.
77. H. Kageyama, The *Origin of Automobile*, Sankai-do Press, Tokyo, Japan, 1999.
78. M. Segawa et al., A study of vehicle stability control by steer by wire system, *Proc. AVEC'2000*, Ann Arbor, MI, 2000.
79. S. Nakano et al., Steering control strategies for the steer-by-wire system, *Trans. JSAE*, 31(2), 2000.
80. K. Sakai et al., A study on control performance of vehicle using electric power steering (EPS) with low torsional rigidity, *Proceedings of Trans-Log JSME*, No 01–36, 2001.
81. S. Nakano et al., A study of vehicle stability control system with active front-wheel angle control, *Trans. JSAE*, 32(1), 2001.
82. T. Matsunaga, T. Tanaka, and S. Nishimura., Analysis of self-excited vibration in hydraulic power steering system: Prevention against vibration by supply line, SAE Paper No. 2001–01–0488, 2001.
83. R. Isermann et al., Fault-tolerant drive-by-wire system, *IEEE Contr. Syst. Mag.*, 22(5), 64–81, 2002.
84. D. Odenthal, How to make steer-by-wire like power steering, *Proceedings of 15th IFAC*, Barcelona, Spain, 2002.
85. N. Inoue et al., Safety evaluation for an electric power steering system, *Trans. JSAE*, 34(4), 2003.
86. Se-Wook et al., The development of an advanced control method for the steer-by-wire system to improve the vehicle maneuverability and stability, SAE Paper, 2003.

87. S. Motoyama, The possibilities of steer-by-wire on vehicle dynamics, *J. JSAE*, 57(2), 39–43, 2003.
88. R. W. Phillips., Self-excited vibration in hydraulic steering racks, SAE Paper No. 2003–01–0580, 2003.
89. R.S. Sharp and R. Granger., On car steering torque at parking speed, *Proc. Inst. Mech. Eng. Part D J. Automob. Eng*., 217, 2003.
90. T. Inoue, M.Q. Dao, and K.-Z. Liu., Development of an auto-parking system with physical limitations, *SICE Annual Conference Program and Abstracts*, Vol. 2004, No.8, 2004.
91. P. Raksincharoensak et al., Front steering control of steer-by-wire based on four-wheel-steering theory, *Trans. JSAE*, 36(2), 2005.
92. N. Zhang and M. Wang, Dynamic modeling of hydraulic power steering system with variable ratio rack and pinion gear, *JSME Int. J. Ser. C Mech. Syst., Mach. Elem. Manufact*., 48(2), 2005.
93. B. Gao, K. Sanada, and K. Furihata, A study on modeling hydraulic-power-steering for heavy duty vehicles, *JFPS Int. J. Fluid Power Syst*., 1(1), 8–17, 2008.
94. T. Miyoshi, Y. Shimizu, and A. Yoneda, Reduction of cogging torque in DC brushless motors for electric power steering, *J. Mech. Syst. Transp Logist*., 1(1), 2008.

26 Structural and Dynamic Problems in Car Body Design

Giovanni Belingardi and Massimiliano Avalle

CONTENTS

26.1 Introduction 943
26.2 Exterior Look and Passenger Compartment Style 944
26.3 Aerodynamics 944
26.4 Bending and Torsion Stiffness 945
26.5 Dynamic Behavior and Modal Analysis of Car Body 951
26.6 Passive Safety 957
26.7 Fatigue Endurance 959
References 963

26.1 INTRODUCTION

The body of a vehicle is a very complex structure. Usually, it consists of thin-walled beams, characterized by an appropriate section shape, and thin shells properly drawn in order to obtain the selected body shape. Figure 26.1 shows an example of a recent car body: body frame structure, panels, lateral doors, back door, and motor compartment hood are clearly visible. Steel in sheets, with a relatively large, and further increasing, variety of steel types available today, is a much utilized material for car body construction. Recently, other materials, both metallic and plastic, are also widely used, with an increasing percentage, mainly due to the lightweight requirements for new vehicles.

The body structure is designed to play a number of different functions, of great relevance for the product quality, that is, for those vehicle characteristics that a customer will consider carefully when he/she decides to buy a vehicle.

In particular, one has to consider the following:

- The room, the exterior look, and the passenger compartment style.
- The aerodynamic shape and performance have a large influence on the fuel consumption (and therefore on the operation costs).
- The bending and torsion stiffness of the vehicle structure have a large influence on handling, on the riding comfort, on the fatigue strength, and even on the entity of squeaks and rattles.
- The anchoring structure for mounting and/or connecting all the vehicle mechanical subassemblies and devices: the internal combustion engine, the gear box, and the motion transmission line, the vehicle suspensions and the steering system, the brake system and the driver pedals, the exhaust pipe, and so on.
- The reliability, with particular reference not only to the fatigue endurance but also to ageing phenomena, that is, the progressive decrease in other quality characteristics with time and use.
- The acoustic and vibration comfort (commonly called NVH or noise, vibration, and harshness) with particular reference to the passenger compartment and with respect to the main noise sources related to the internal combustion engine, the gear box, and the tire rolling.

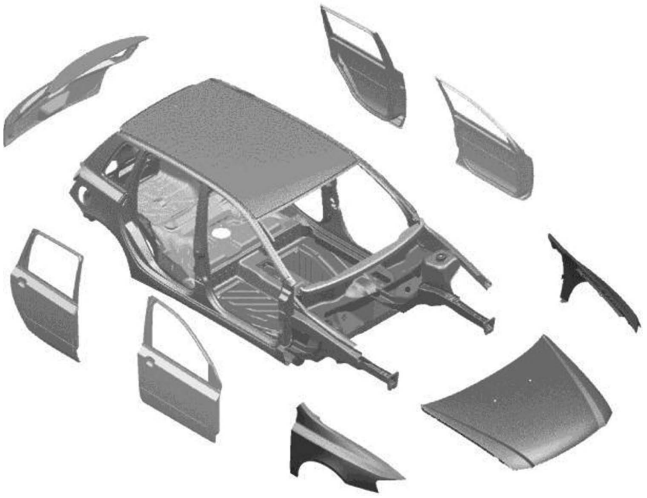


FIGURE 26.1 Body of a recent car with five doors.

- The thermal/climatic comfort inside the passenger compartment.
- The comfort for the vehicle driver and the passengers, in terms of both accessibility and seating position during the ride; for the vehicle driver position, there are some other additional requirements related to the easy accessibility and maneuverability of the commands.
- The passenger safety in case of accident.

These points will be briefly examined in the following paragraphs with particular attention to the more recent solutions and design approaches.

26.2 EXTERIOR LOOK AND PASSENGER COMPARTMENT STYLE

Vehicle style is the capital, although not unique, factor for the sell success of a vehicle. Vehicle style has, on one side, details distinctive of the car manufacturer that are maintained over the years and, on the other side, specific characteristics linked to the customer preference evolution, taking into account the differences encountered in the different markets. Nowadays, it is evident that there is a tendency toward vehicle personalization as a distinctive mark of its owner. Personalization is obtained through the careful choice of a number of aesthetic details that differentiate the available versions of the same vehicle. Moreover, there are the so-called niche models, that is, vehicles produced in small/medium series with an exterior look remarkably differentiated with respect to the mass production vehicles (e.g., the mono-volume vehicles that, starting with the Renault Espace, have progressively gained relevant market portions, in comparison with the more traditional sedans of the two- or three-volume types).

Nowadays, the passenger compartment has also definitely gained importance in vehicle aesthetic characterization and increasing attention is paid to the style of the seats, dashboard, instrument set, steering wheel, door inward panels, and other interior dressing panels. Finally, it is worth noting that a little attention is also paid to the engine compartment dressing; even in this compartment, nowadays, an orderly and pleasant disposition of different devices, electric cables, and tubes is expected.

26.3 AERODYNAMICS

The dramatic energy crisis of the 1970s contributed to point out the role of the aerodynamic performance of the exterior shape of the vehicle with respect to the fuel consumption while riding [1,2]. Since that date, systematic studies about the aerodynamic profile of the external shape of the

body have been developed, particularly aimed to reduce the longitudinal aerodynamic coefficient C_x. Refinement of the analysis techniques allowed to study in detail the dependence of all the six aerodynamic coefficients on the vehicle characteristic motion angles with respect to the air flow direction. From this point of view, particular attention has been paid to the effect of the lateral aerodynamic flow on the driving stability through specific studies on the sensibility of the vehicle trajectory to the lateral wind.

The external shape has been progressively modified over the years, and optimal parameters have been identified, such as, for example, the windscreen angle and the rear truncation shape, both for two-volume and the three-volume vehicles. Optimized shape and position for aerodynamic profiles have been researched: Front aerodynamic dam has been added to improve the air flow through the water cooler of the internal combustion engine, appropriate position has been determined for the intake vent to improve the passenger compartment air flow, and rear spoilers have been added [3]. Moreover, the external shape of the vehicle deck has been specifically studied in order to have appropriate air flow under the vehicle with a little lift aerodynamic effect. The type and shapes of the external joints of the body structure parts have been deeply modified: The drop channels, which were typical of the automobiles up to the 1970s, progressively disappeared and the modern roof is smooth; the window sealant contours, especially in the case of the front windscreen, have been redesigned in order to avoid the detachment of the air flow and the formation of vortexes; and the external shape of the lateral mirrors and door handles have been modified with the same scope for aerodynamic performance optimization [3,4].

Furthermore, the aerodynamic design of the vehicle body has led to important improvement in the air flow inside the passenger compartment. The inlet and outlet vents have been properly positioned in order to obtain more favorable air mass flow. Careful studies of air flow inside the passenger compartment have been conducted in order to improve both the internal climatization, toward a good uniformity of the temperature, and the windscreen wipe, against mist.

Research in the field of aerodynamics has developed in two main directions: On one side, the experimental test through the design and development of aerodynamic wind tunnels equipped with specific instruments and sophisticated measurement devices for the analysis of aerodynamic flow; on the other hand, the numerical simulation through the development of computational codes, based on the fluid domain discretization by a specialized finite element approach, that permits a detailed analysis of the aerodynamic behavior of the vehicle profile. The use of the numerical approach requires very large computing machines because it is necessary to discretize, with good quality mesh, details of a vehicle's external shape, with good refinement in the air volume in the neighborhood of the body, and with sufficient detail of other air mass inside the wind tunnel.

The persevering comparison of the measured experimental results, not only the global values expressed by the six aerodynamic coefficients characteristic of the vehicle body but also the values of the specific localized phenomena, with the numerical simulation results, has permitted the development of a good mathematical model. Furthermore, through this analysis, the profile designer has gained better knowledge of the influence of each shape detail on the vehicle global behaviour.

26.4 BENDING AND TORSION STIFFNESS

The vehicle structure has its own values of the bending and torsion stiffness, and both are calculated on the basis of the displacements, measured at one of the axles with respect to the other one that is completely clamped, due to vertical forces applied at the wheel centers. These stiffness values are the consequence of the structure configuration that has been chosen for the base frame and for the deck. These values increase when the windshield and the window glasses are applied, due to their effect as shear panels, and when the whole vehicle is considered. Data on the body stiffnesses are reported in Table 26.1.

TABLE 26.1
Comparison of Torsion Stiffness K_t and Bending Stiffness K_f Values for European Vehicles of the Same Market Segment

	Torsion Stiffness, K_t (kN-m/rad)			Bending Stiffness, K_f (kN/m)			Glass Contribution to Stiffness (%)		Torsion Stiffness to Weight Ratio
Vehicle Model	Body-in-White	Body w. Glasses	Complete Body	Body-in-White	Body w. Glasses	Complete Body	K_t	K_f	Body-in-White (m/rad)
Fiat Punto MY94 5 doors	573	701	796	6.3	6.4	6.7	22	2	302
Renault Clio MY90 5 doors	540	740	770				37		355
Fiat Uno MY90 5 doors	342	404	478	4.3	4.4	4.8	18	3	225
Ford Fiesta MY89 5 doors	420	515	665	3.5	3.6	3.8	23	1	258
Lancia Y MY85	445	556	678	6.4	6.8	7.6	25	6	295
Citroën AX MY87	455	635	690	4.6	5.0	5.7	40	10	358
Peugeot 205 3 doors	390	500	588	3.2	3.6	3.8	28	11	275
Fiat Punto MY94 3 doors	578	728	834	6.0	6.2	6.3	26	4	312
Fiat Uno MY90 3 doors	336	423	486	4.4	4.6	4.8	26	3	224
Peugeot 106 3 doors	567	730	820	5.7	5.9	6.4	29	4	
Opel Corsa 3 doors	410	540	690	4.9	5.0	5.1	32	2	
Nissan Micra 5 doors	300	370	510	4.8	4.9		23	2	
VW Polo MY94	550	660	790	4.1	4.2		20	2	
VW Polo MY82 3 doors	360	380	435	4.9	5.0	5.2	6	2	242
Renault 5 5 doors	340	480	530	4.2	4.5	4.8	41	6	243
Fiat Palio 2V 5 doors	428	579	683	5.8	6.0	6.2	35	3	222
Lancia Y10	583	735	835	6.8	8.9	7.2	26	31	

In order to obtain the highest values of both stiffnesses, the following aspects are very important:

- The layout of the longitudinal and transverse beams (see Figure 26.2a)
- The wall thickness
- The shape of the beam section (that very often has an internal wall in order to increase the section moment of inertia and to oppose the wall buckling)
- The design of the structural multiple joints (they are put in evidence in Figure 26.2b; this last aspect is of great importance)

Furthermore, it is well known that the stiffness improvement is given by the panels that are joined to the beam skeleton by welding. Commonly, electric spot welding is used; recently, laser welding has also been employed, both continuous and dash, with a slight although appreciable increment in the structure stiffness [5–12]. More recently, there are also some adhesive joint applications; this type of joint requires proper design of the joint itself to take full advantage of technology [13–20].

The panels (see Figure 26.3), according to the structural scheme originally proposed by Pawlowski [21] and still in use to develop basic structural considerations, are submitted to tensile, compression, and shear loads. Their relevant shear properties contribute in a decisive manner to increase the structure stiffness, substantially locking the deformation modes with rigid rotation around the hinges, typical of the articulated parallelograms that could take place in the body frame structure due to the finite compliance of the joints [22–24].

The care in the design of structural details, together with a refined analysis by the finite element technique, led to a relevant increase in the body stiffness performance even without an increase in the structure weight.

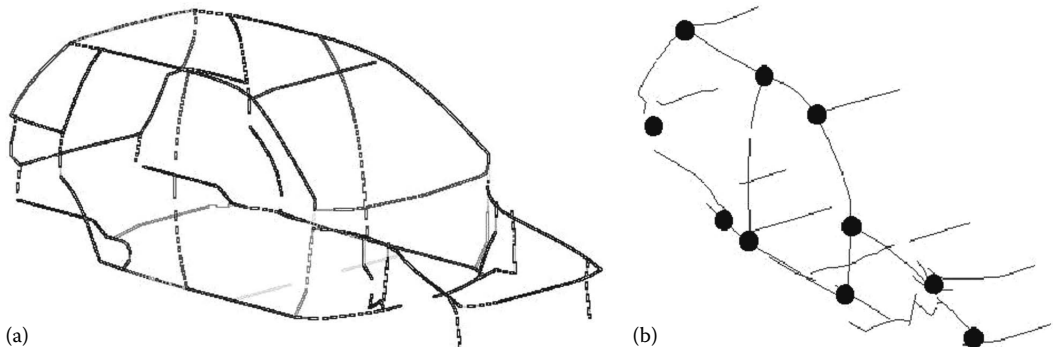


FIGURE 26.2 Layout of the beams in the body structure (a) and structural nodes (b).

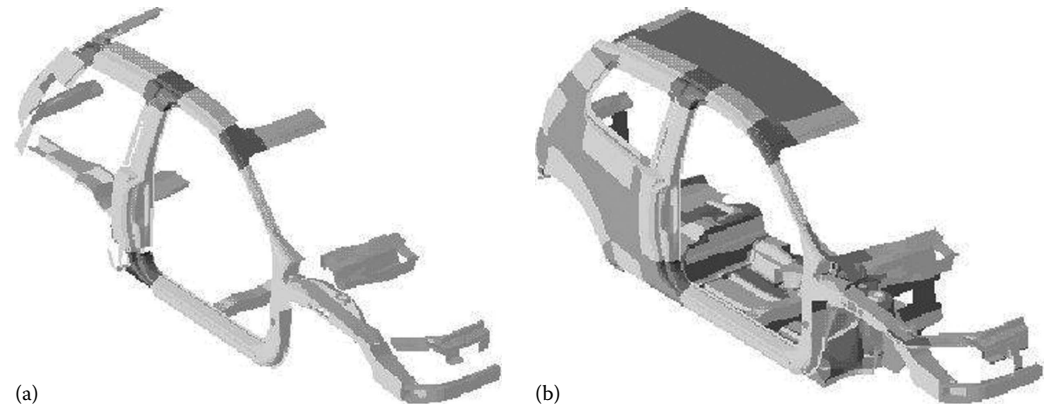


FIGURE 26.3 Structural model of the vehicle body beams (a) and model with fixed panels (b).

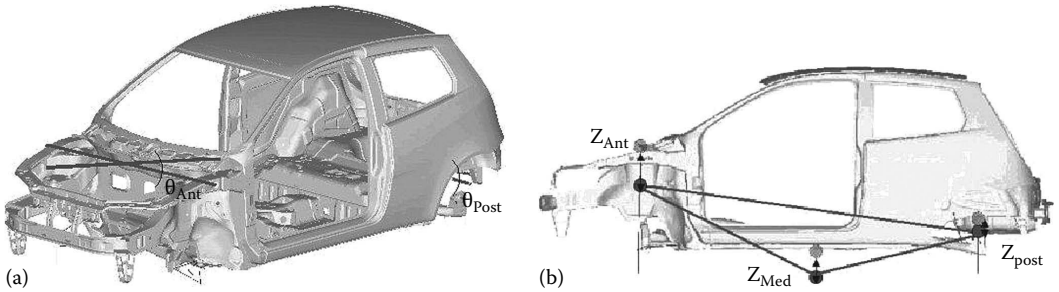


FIGURE 26.4 Evaluation of the torsion (a) and bending (b) stiffness of a car body.

Figure 26.4 shows the scheme of loads and constraints used for the evaluation of the torsion stiffness (Figure 26.4a) and bending stiffness (Figure 26.4b) of a car body. Table 26.1 collects data of stiffness for a number of vehicles of the same market segment. The vehicle list is enriched by two relevant information: the model year and the number of doors. By comparison of the values of the new model with respect to the previous one (see, e.g., Fiat Punto versus Fiat Uno), it is possible to have an appraisal of the evolution of these indexes with time toward higher values, as a consequence of appropriate design choices. By comparison of the values of the five-door version with the three-door version of the same model (see, e.g., Fiat Punto), it is possible to have an appraisal of the effect of this architectural difference.

In Table 26.1, stiffness values are reported specifying the contribution of the body in white, the body with glasses, and of the complete vehicle. The contribution of the glasses (and in particular of the windscreen) is very important for torsion stiffness; in most cases, it ranges between 20% and 40%.

High-performance vehicle and sports cars, due to major requirements in drive precision, are characterized by higher values of their torsion stiffness; typical values for the body in white structure are in the range of 850–1700 kN-m/rad. Lamborghini, for the Gallardo coupé model, gives the torsion stiffness value of 1600 kN-m/rad.

In many cases, these high-performance cars are in the coupé version architecture. This architecture reduces and very often eliminates the contribution of the upper part of the body structure to the vehicle torsion stiffness. Therefore, the lower part of the body structure has longitudinal and transverse beams of larger sections, in order to recover the desired value of the stiffness.

In recent years, a tendency toward the use of frame structure can be noticed [7–11,25–28]. The frame structure is substantially based on the use of prismatic beams of the desired cross section, connected by three-dimensional joints, often made by casting. Prismatic beams are generally made by extrusion of aluminum alloys and the desired curved axis line is obtained by cold stamping or hydroforming [28–36]. By using beams of different lengths but of the same cross section and with the same three-dimensional joints, it is possible to make car bodies of different dimensions, even with small production quantity (the niche vehicles); thus, vehicle personalization is feasible not only with respect to the appearance of the vehicle but also with respect to the vehicle dimensions.

Three other points are to be noted in the recent evolution of the car body construction [37,38]:

- The use of steel types different from the traditional low-carbon mild steel that is preferred by manufacturing people because of its particular suitability for deep drawing
- The use of material types in sheet different from steel to make structural parts
- The use of joining technologies different from the usual electric spot welding for the assemblage

The traditional steel types have had a significant evolution [37] in order to obtain more ductility, aimed to increase deep drawing formability, and oxidation resistance by definition of specific surface passivating treatments and painting technology.

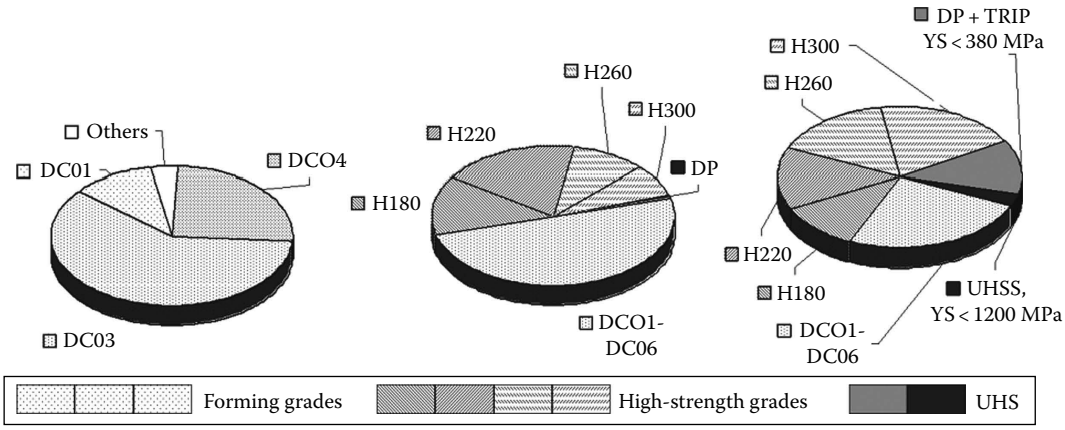


FIGURE 26.5 Past and present situation and future trends in the steel grades' utilization for car body construction (DC01-DC06, EN 10030; H130-H300, high-strength steels, yield strength 130–300 MPa; UHSS, ultrahigh strength steel; DP, double phase steel; TRIP, transformation-induced plasticity steel; YS, yield strength).

New steel families of very high performance have been developed, besides the so-called HS (high-strength) steels; nowadays, UHS (ultra high strength) steel and dual phase steel with bake hardening characteristics are also available [29,37–43]. Figure 26.5 illustrates the use of different grades of steel in car body construction 25 years ago, at present, and in the near future [29,37].

The steel types characterized by high-strength grades exhibit higher values of the yield and ultimate strengths but smaller ductility and fracture strain with respect to the traditional low carbon steel. The use of this type of high-strength steel allows for reduction in the wall thickness of the structural members formed from sheets and thus for relevant reduction in the weight of the car body structure. Of course, the reduction in the wall thickness has a negative influence both on the global torsion and bending stiffness and on the local buckling values; the desired values must be recovered by an appropriate design of the beam section shapes.

Once again, in order to obtain lightweight car body structures, aluminum solutions have been developed; in this perspective, the Ferrari Modena solution and the solution, known as the ASF or Audi Space Frame, developed by Audi in cooperation with Alcan and adopted for the Audi 8 and then for the Audi 2 models, and the Jaguar XJ8 have to be mentioned. Table 26.2 lists the most notable space-frame-based vehicles, together with information on materials and assembly methods.

Aluminum types for car body are in the 5xxx and 6xxx range [34,44–50].

The Audi 8 first generation was the first model produced; it had a relatively small quantity of production per year. Audi 2 followed: it had a relatively large quantity of vehicle production. Therefore, it was decided that Audi 2 had a completely new design solution that took advantage of the experience gained from the previous model but also introduced deep design and manufacturing changes in the perspective of production of a larger quantity.

The ASF of Audi A8 was revolutionary, but many improvements came with the A2 frame [37]. The original A8 space frame made extensive use of simple, cold-extruded aluminum tubes connected to each other with complex rigid cast joints by MIG welding and other mechanical joining methods, such as punch riveting.

The A2 space frame is made with rolled aluminum profiles and simpler cast joints, assembled by laser welds and adhesives. Cost savings has been obtained through a drastic reduction in the number of parts (from 334 to 225) and with extensive use of automation for assembly operations.

Some particular parts, at the moment mainly lateral doors [51] and seat structures, made of magnesium are also to be mentioned as interesting experiences of using light metal other than aluminum for structural parts.

TABLE 26.2
Main Up-to-Date Marketed Space-Frame-Based Vehicles

Brand/Model	Year	Structure Type	Materials	Assembling Methods
Ford AIV	1998	AVT[a], stamped sheet	AA5754-0	Weld bonded
Audi A8	1994	ASF[b], cold extruded aluminum tubing	Aluminum alloys	Rigid cast joints, MIG welding, punch riveting
Audi A2	1999	ASF[b], aluminum profiles	Aluminum alloys	Simpler cast joints, laser welds, and adhesives
Ferrari 360 Modena	1999	Aluminum ALCOA space frame (153 kg)	6008-series Al alloy (mainly 6008)	Sand cast structural joints, MIG welding
Volvo XC90	2002	High-strength steel frame	HSS, EHSS, UHSS	Welding (various) and brazing, adhesives, fasteners
Jaguar XJ8	2004	Ultrastiff stamped aluminum monocoque	Aluminum alloys	Bonded and riveted
Chevrolet Corvette Z06	2006	DANA space frame (129 kg)	Aluminum alloys	Self-piercing rivets (236), laser welding (14 m)

[a] Aluminum vehicle technology (ALCAN).

[b] Audi space frame.

Finally, some experience in the making of relevant parts of the car body structure, such as the deck, with fiber-reinforced plastic material has to be mentioned [52]. Composite materials are of great interest as they have very high strength to density ratio and therefore allow for very light structures [53–56]. Due to cost reasons, which is still rather high for composite materials in comparison with steel, glass fibers have been mainly used, although carbon fibers have been also considered. The possibility of sandwich structures made by thin skins of plastic matrix composite material and polymeric foam or honeycomb core can be the desired compromise [25,53,57–60].

The adoption of materials other than metals requires careful consideration and satisfying solutions for dismounting and material recycling at the end of the vehicle's life.

In the field of the part-joining technologies, beside the classical spot welding and, in some cases, the bolting of metal sheets, applications of great relevance have been developed by laser welding and structural adhesives [5–7,12,18–20,61–64]. While solutions with classical riveting, self-perforating riveting [64–68], and clinching have also been experimented with, it is clear that joining technology other than welding, and in particular the solution with structural adhesive, is needed when materials other than metals are used.

Great advantages come, not only for each structural component but even for the whole car body, from the extensive use of adhesives [64,69,70]. Adhesive joining solutions have already been developed for some high-class European cars (BMW, Mercedes, and Jaguar [67,68]). Adhesives offer many improvement possibilities to the car body construction. Since the adhesive joint is continuous instead of a series of spots, stiffness characteristics of bonded components are higher. Furthermore, adhesives act as a sealant, providing the advantage of good protection against seepage and corrosion.

The adhesive solution offers potential for great improvement in the overall bending and torsion stiffness of the car body [71–74]. It has been shown that bonded crash columns can absorb and dissipate much more energy than an equivalent spot-welded column [75], provided there is a suitable design for the bonded column: For example, the usual half-shell with flanges [13,20] to be spot welded should be substituted with pieces without flanges. Furthermore, adhesives have high damping properties and help in improving the NVH characteristics of the car [76,77].

One of the drawbacks of the use of adhesives is the long manufacture cycle time required to obtain a sufficient level of polymerization of the adhesive [78]. For that reason, it is necessary to

use additional joining devices to hold the sheet in place until the adhesive is able to develop its full strength. Another possible concern is the difficulty in preparing and putting in place the sheets for adhesive joining, because this asks for higher precision with respect to spot-welding technology. However, this problem can be solved in the concurrent design process of the body shell.

Since it is not ideal to keep together the sheets to be bonded with spot welds, alternative joining methods have to be found. Mechanical fastening can be used. Traditional riveting is not likely to be used for several different reasons (including aesthetic appearance) but punch riveting [79,80] or clinching [81] is effective for solving this problem. Both methods can be easily automated and are also cost-effective solutions [56,82].

26.5 DYNAMIC BEHAVIOR AND MODAL ANALYSIS OF CAR BODY

The car body is characterized by its static torsion and bending stiffness. These values are important indexes of the car ride comfort and handling because the car body as a whole can be considered as one torsion and one bending spring placed in series with respect to the front and rear suspensions.

Beside these static characteristics, the vibration and acoustic characteristics are also of relevance for a car body. Figure 26.6 shows a schematic subdivision of the frequency range of interest for car body comfort, distinguishing ride vibrations from harshness and noise.

The dynamic behavior of a vehicle is strongly influenced by the global and local dynamic stiffness, structural damping, and vehicle mass [83–85]. For the acoustic performance of the passenger compartment, the quantity and layout of the phono-absorbing and phono-isolating materials are also of importance.

Vibration and acoustic phenomena are determined by the dynamic excitations that come from different sources (see Figure 26.7) such as the motor engine, the gear box and the drive line, the tire rolling, and the aerodynamic air flow. Each of these sources is dominant in a certain range of frequency. In order to study the noise level inside the passenger compartment, it is important to consider and identify two different transmission paths: the structure path and the air path.

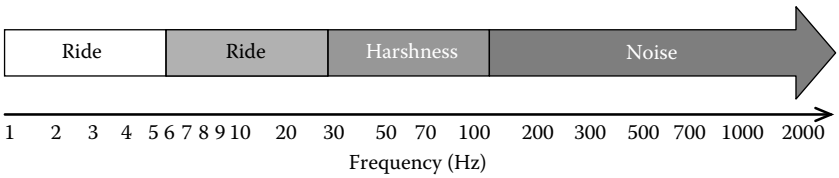


FIGURE 26.6 Characteristic frequency ranges for vehicle comfort.

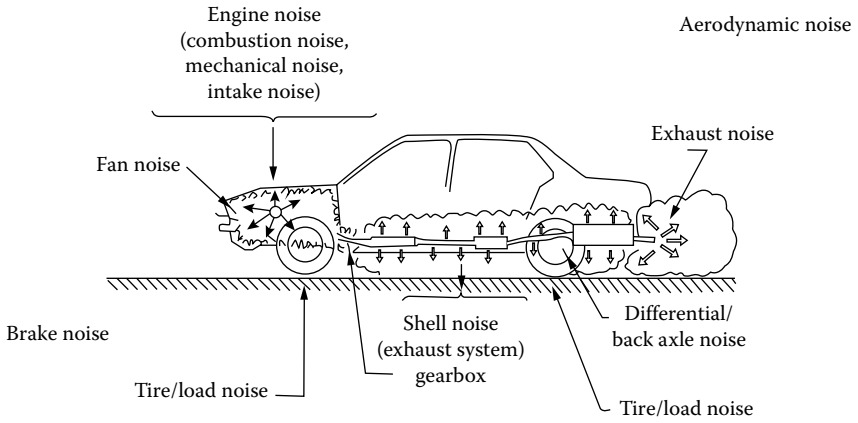


FIGURE 26.7 Noise and vibration sources.

Let us consider as an example the vibration of a motor engine [84–89]. When the engine is running, it vibrates at frequencies that are multiples of the rotational speed of the engine itself. The noise of the engine is heard directly from its walls; this noise has a particular intensity level when the excitation frequency is equal to one of the engine structure eigenfrequency. The acoustic waves that are present in the engine compartment pass through the body structure and enter the passenger compartment (the so-called acoustic "transparency" is a measure of this phenomena). This type of noise is usually called "air-borne" noise (as it is transmitted through the air path) and is generally predominant at the higher frequencies (>500 Hz).

The car body vibration is excited also by the cyclic forces that are applied at the engine mechanical links. Vibration energy is thus transmitted through the solid structure to the human–machine interface points (such as the steering wheel, the seats, the pedals, the feet places on the deck, etc.). Vibration energy is also transmitted to the passenger compartment panels that begin to vibrate themselves, producing further internal noise. This type of noise is usually called "structure-borne" noise (as it is transmitted through the structure path) and is generally predominant at lower frequencies (<200 Hz).

The performance of the car body in the noise transmission perspective is quantified through the acoustic transfer function *P/F*, where *P* is the noise at one point inside the passenger compartment—usually, microphones are placed corresponding to one of the ears of the driver or to that of one of the passengers—while *F* is the exciting force applied to one point of the body structure. This transfer function is clearly a frequency function (Figure 26.8).

The overall value of the noise heard by the ear, caused by the motor engine, can be calculated by the summation of the contribution of all the engine mechanical link points using the following equation:

$$P = \sum_{n=1}^{m} F_n \times \left(\frac{P}{F}\right)_n \cong \sum_{n=1}^{m} (K_n \times X_{mtp,n}) \times \left(\frac{P}{F}\right)_n \tag{26.1}$$

where *m* is the number (one for each loading direction) of these engine link points to the car body.

By means of this equation, it is possible, once the objective for the overall internal noise *P* has been fixed, to get partial objectives for the subsystems, in particular for the elastic characteristic curve *K* of the engine mounts, for the displacements X_{mtp} of the motor engine group, and for the *P/F* acoustic

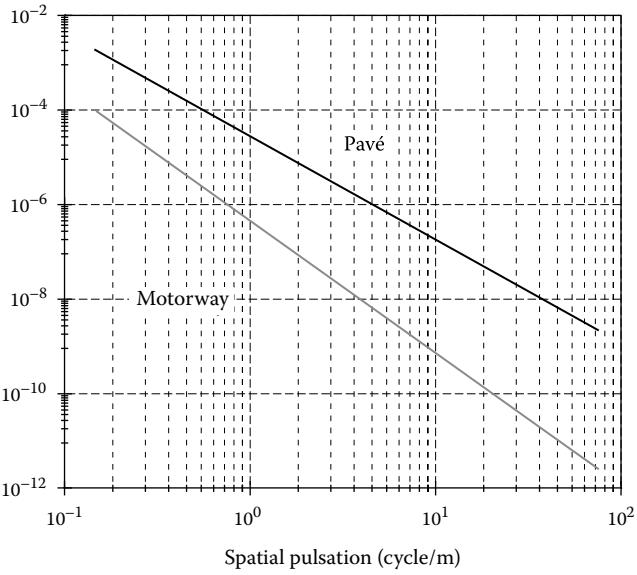


FIGURE 26.8 Excitation from the road surface irregularities: comparison of the power spectra characteristic of a motorway and a pavé.

transfer function of the body. Obviously, the displacement X_{body} of the link point of the engine mount to the body must be negligible in comparison with the displacement X_{mtp} of the motor engine such that it can be neglected in Equation 26.1. A sample *P/F* transfer function is shown in Figure 26.12.

With the scope of quantifying the dynamic response of a structure, the acceleration over force transfer function that is easily obtained from experimental tests is generally used. When the input (excitation) point and direction coincide with the output (acceleration) point and direction, this transfer function is called inertance function:

$$I_k = \frac{\ddot{u}_k(\omega)}{f_k} = \frac{-\omega^2 u_k(\omega)}{f_k} = \sum_{i=1}^{n} \frac{\omega^2 \varphi_{ki}^2}{-\omega^2 m_i + i\omega b_i + k_i} \tag{26.2}$$

$$K_{dyn,k} = \frac{f_k}{u} \rightarrow K_{dyn,k} = -\frac{\omega^2}{I_k} \tag{26.3}$$

where

I_k is the inertance function

$K_{dyn,k}$ is the dynamic stiffness function for the load application point k

In Equations 26.2 and 26.3, according to the usual modal analysis symbols,

ω is the angular pulsation considered

i is the eigenmode index

m_i, b_i, and k_i are the modal mass, the modal damping, and the modal stiffness, respectively, associated to the i eigenmode

φ_k is the k component of the i eigenmode

Equation 26.3 puts in evidence that the higher the inertance value, the lower the structure dynamic stiffness. In Figure 26.9, some examples of the inertance curve are shown and, for the sake of comparison, three curves for the constant stiffness case are also drawn.

The dynamic stiffness of a car body is the result of a number of structural vibration modes that are present all together. These structural modes can be subdivided in two groups: the global modes

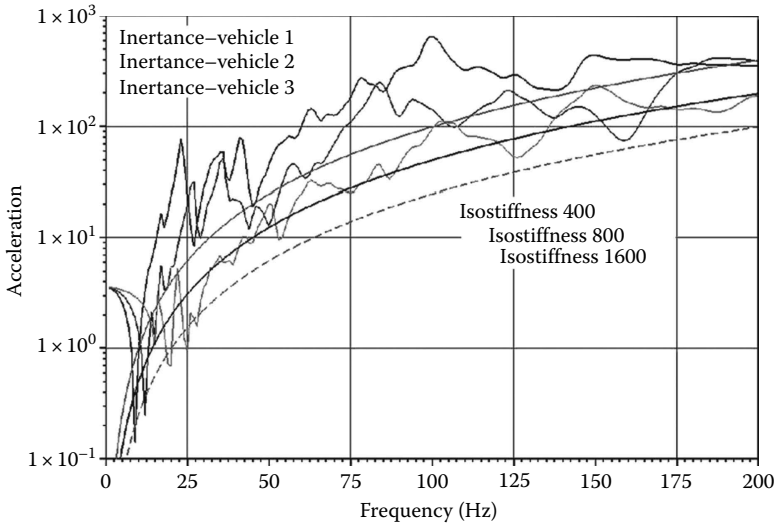


FIGURE 26.9 Inertance curves and constant stiffness curves versus frequency.

and the local modes. The first group includes the modes that involve the whole structure with vibration of some amplitude; the second group includes the modes that involve only one substructure, more or less localized; while the rest of the structure is nearly not vibrating.

The global modes are at the lower frequencies, generally lower than 50 Hz. They are easily detected and classified as torsion modes, bending modes, and waving modes.

For each of these modes, it is further possible to distinguish if the modal shape has its larger displacement at the front part or at the rear part of the car body structure. It is well known that the modal shapes are independent from the entity of the applied loads; therefore, the dynamic deformed shape of the car body, as a consequence of the superposition of its global modes and of a mass distributed in a more or less uniform way, could be quite different from the static deformed shape.

In Figure 26.10, modal deformation shapes of a body structure for some global modes are shown; the corresponding eigenfrequencies are also reported. The car body exhibits two torsion modes: one is of the global type and the other one has larger deformation motion at the front part of the structure.

As already mentioned, the mass distribution is relevant for the eigenfrequencies of the global modes. Therefore, one has to consider that when all the other devices, covers, and internal dressing elements are applied to the car body, the mass values are increased, and somehow its distribution modified, without relevant modification in the structure stiffness and, therefore, without changes in the static stiffness values. In contrast, the dynamic response changes and the eigenfrequencies of the global modes become lower [86].

Global modes are very important from several points of view: first of all, the coupling of the global modes with the vehicle suspension modes and, eventually, with the other supporting frames that are connected to the car body through more or less rigid links. It is clear that it is of primary importance that these link points move as little as possible during the dynamic vibration of the structure. Thus, the dynamic vibration of the car body can influence the drive quality of the vehicle itself in a remarkable manner, in terms of both handling precision and comfort.

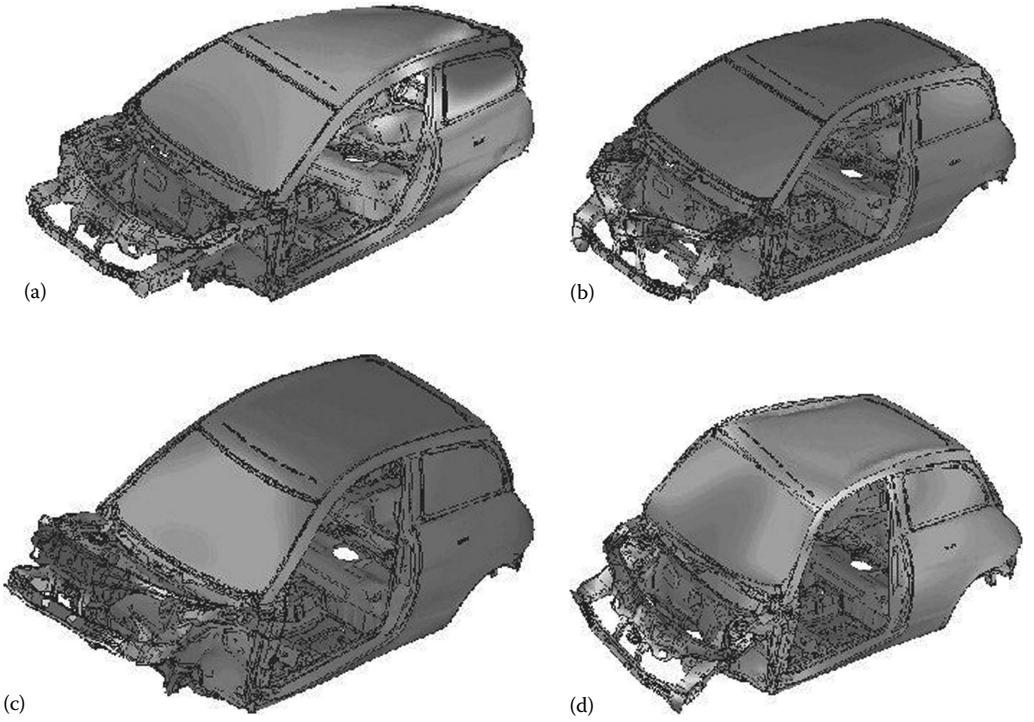


FIGURE 26.10 First four modal shapes of a car body with glasses but without movable parts: (a) global torsion mode, (b) front structure waving mode, (c) front bending mode, and (d) second front torsion mode.

Another relevant point is the possibility that the structure global eigenfrequencies are in resonance with other vibration frequencies, such as the excitation frequency of the motor engine at idle or the eigenfrequencies of the steering wheel–steering column–steering mechanism. In both cases, the frequency coupling results in a remarkable amplification of vibration and noise in the passenger compartment. In these cases, it is possible to obtain a solution to the problem by the so-called structural modification procedure [90]. This procedure is based on the modal model obtained by the experimental modal analysis. A proper modification of the system modal mass and/or the system modal stiffness is "virtually" made in order to move the eigenfrequency and to decouple the resonant modes. From the modified modal model, it is then necessary to go back to the real model and to determine the consequent modifications of the system parameters. It is not easy to give an indication of the distance between frequencies to guarantee their decoupling, as it is influenced also by the value of the modal damping.

Let us move to the vibro-acoustic phenomena. Besides the body structure eigenmodes in this perspective, the volume of the air cavity inside the passenger compartment must be carefully considered. The air cavity has its own dynamic behavior characterized by eigenmodes and eigenfrequencies that are present and even coupled with those of the body structure. The cavity properties are very important to the evaluation of the noise level heard by the passengers.

The air volume has its own dynamic behavior that is obviously different from the structural behavior, as it is constituted by fluid. In the analysis phase, it is possible to use an approach that is similar to the modal analysis already addressed for the structure vibration analysis. As for the structure, it is possible to evaluate the overall response as summation of a number of contributes due to the different modal deformations; in a similar way, it is possible to evaluate the acoustic response of the passenger compartment cavity as the summation of its pressure eigenmodes.

Table 26.3 collects typical values of the three lower eigenfrequencies of the passenger compartment cavity modes, distinguishing between vehicles of different market segments.

For a cavity of simple regular geometry, it is possible to calculate the eigenfrequencies of the cavity modes easily through an analytic approach; they are inversely proportional to the time taken by an acoustic (pressure) wave to travel back and forth in the cavity in one considered direction. The pressure wave propagation velocity in air is the sound speed (344 m/s in normal conditions, i.e., 293 K, 101,325 Pa).

The cavity modes (see, e.g., Figure 26.11), similarly to the structural modes, have node points, where the pressure dynamic variation is null, just as in the vibration modes where displacements were null. If the driver or the passengers have their ears near these points, they will hear little noise component at the corresponding frequency, with a beneficial effect on the perceived noise level and consequently an increment in the internal comfort.

In the same manner, if a noise source (e.g., a vibrating panel) is placed near one of these acoustic nodes, it will not be able to excite the corresponding cavity mode and the internal noise will not suffer due to its contribution. On the contrary, if acoustic sources or the passengers' ears are close to the points of cavity mode maximum amplitude, internal noise results to be amplified with a detrimental effect on the perceived acoustic comfort.

TABLE 26.3
Typical Values of the Lower Cavity Eigenfrequencies for Different Vehicle Segments

	Typical Frequencies of Cavity Modes		
	Small Car Segments (Hz)	**Compact Car Segments (Hz)**	**High-Class Car Segments (Hz)**
First longitudinal mode	60–70	55–65	45–55
Longitudinal mode	115–130	100–115	85–100
First transverse mode	130–145	120–135	110–130

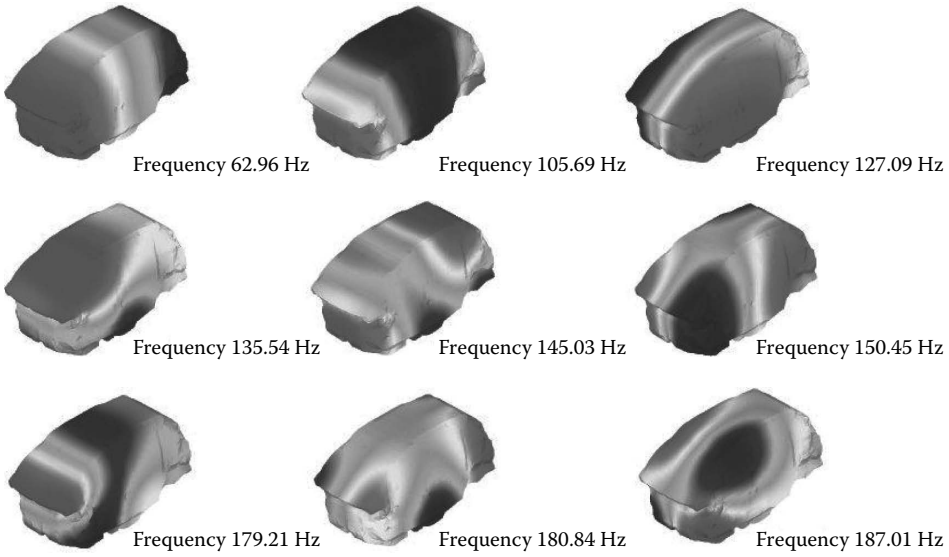


FIGURE 26.11 First nine cavity modal shapes and the related eigenfrequencies.

The air motion in the passenger compartment and structural vibrations are dynamically coupled. When the structure vibrates, the panels that contour the passenger compartment impose their kinematical movement as boundary conditions to the air. In contrast, the local air pressure imposes pressure values as boundary conditions to the dynamic deformation of the structure.

It is possible to apply the modal analysis methodology to the structure and to the cavity coupled [91,92]. Thus, the following equations system could be written as follows:

$$[-\omega^2[M]_s + i\omega[B]_s + [K]_s] \times u(\omega) = F(\omega, t) + [A]^T \times p(\omega) \tag{26.4}$$

$$[-\omega^2[M]_f + i\omega[B]_f + [K]_f] \times p(\omega) = -[A] \times \ddot{u}_n(\omega) \tag{26.5}$$

where

M, B, and K are the mass, damping, and stiffness matrices of the structure (subscript s), respectively, and of the fluid (subscript f)
A is the coupling surface
u is the displacements
p is the pressures
ω is the pulsation

The resulting coupled dynamic problem has the following formulation:

$$\begin{bmatrix} -\omega^2[M]_s + i\omega[B]_s + [K]_s & -[A]^T \\ -\omega^2[A] & -\omega^2[M]_f + i\omega[B]_f + [K]_f \end{bmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix} \tag{26.6}$$

From the solution to this system, both the structural displacement u, taking into account the air motion, and the air pressure p, taking into account the panel motion, can be evaluated. Figure 26.12 shows some results of a research study on the dynamic coupling of structural modes with cavity modes for a vehicle passenger compartment. Figure 26.12 shows a comparison between experimental test results and numerical simulation results.

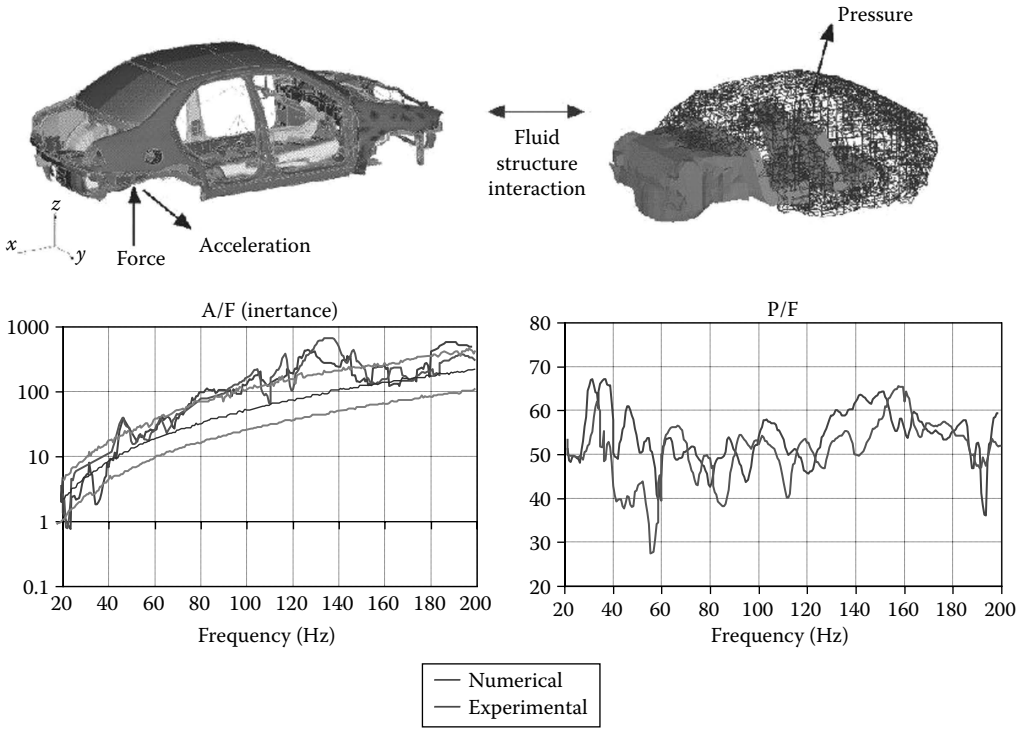


FIGURE 26.12 Structural modes and cavity modes coupling, with the qualitative inertance and pressure responses as a function of frequency.

26.6 PASSIVE SAFETY

Passive safety is the protection that the vehicle structure and the on-board devices (e.g., the safety belts and the air bags) give to the driver and to the passengers in case of an accident, their operation being automatic, that is, does not need commands by the driver for their activation [93,94].

In case of collision with another vehicle or with a fixed obstacle, the kinetic energy of the vehicle must be completely dissipated and the vehicle stopped in a time duration of the order of magnitude of one tenth of a second.

The kinetic energy is absorbed by the plastic deformation of the material of the body structure if it is a metallic material, or by fracture and fragmentation if it is a brittle plastic material such as composite materials with polymeric matrix.

In the perspective of passenger safety, the progressivity of the energy absorption phenomena is fundamental just as the accelerations and the forces to which the different parts of the human body are submitted to do not exceed the tolerability values that are typical of the human body, that is, those limit values over which permanent injuries or even the death of the person is caused.

With this scope, the vehicle body structure (see, e.g., Figure 26.13, where the characteristic components of the front structure of a recent vehicle is shown) is designed in order to get a vehicle deceleration history that is progressive. Figure 26.14 shows the typical history of the deceleration measured on the vehicle deck at the front seats' anchoring points during a frontal crash carried out according to the EuroNCAP testing procedure. In the graph, the different levels of the deceleration that are due to the different parts of the vehicle structure can easily be detected.

According to the EuroNCAP test procedure [95,96], the vehicle impact occurs against a deformable barrier that is intended to reproduce the impact deformation and energy absorption capability

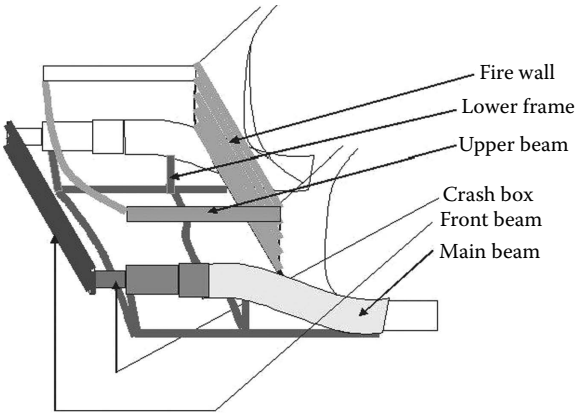


FIGURE 26.13 Characteristic components of the front structure of a recent vehicle.

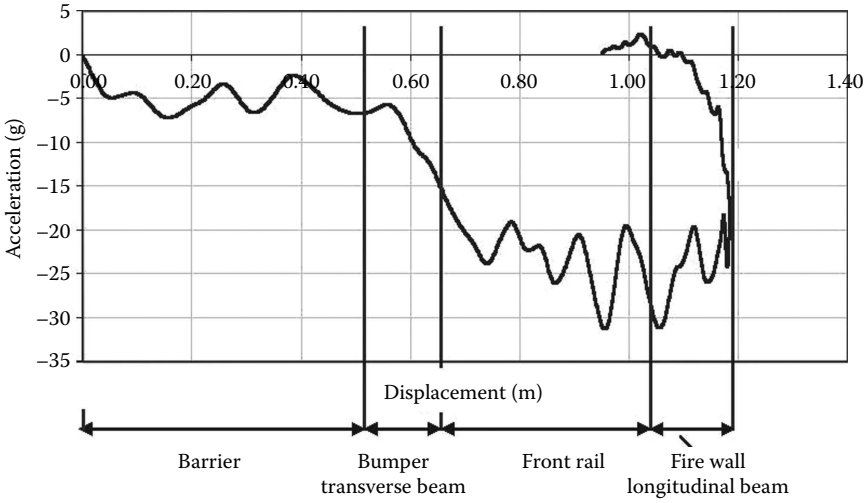


FIGURE 26.14 Typical course of the deceleration during a EuroNCAP crash test.

of the other vehicle when a "car-against-car" impact is considered [97,98]. The initial part of the curve shows the deceleration due to the compression failure of that deformable barrier; then, in the order, the deformation of the front transversal beam; that of the so-called crash boxes; and finally, that of the front structure longitudinal beams follow. Usually, these structure beams are thin-walled boxed beams made of drawn steel sheet and joined by spot welding.

The beam deformation mechanism of plastic axial collapse resulting in the formation of a sequence of folds is the most favorable mechanism that guarantees the absorption of the desired amount of energy and in the meantime satisfies the progressivity requirements [93,94].

The initialization of every new fold determines an increment of the beam axial antagonist force and, therefore, an increment in the vehicle deceleration. This increment is detectable in the deceleration history diagram as a peak. The plastic deformation evolves then through the development of localized plastic hinges in the beam walls. These localized plastic deformations of the metallic material give the desired energy absorption (see Figure 26.15 for typical specific absorbed energy for some structural car body materials).

It is thus evident that the structural behavior of the vehicle body during its collapse is the determinant factor for the entity of the deceleration to which the driver and the passengers are submitted.

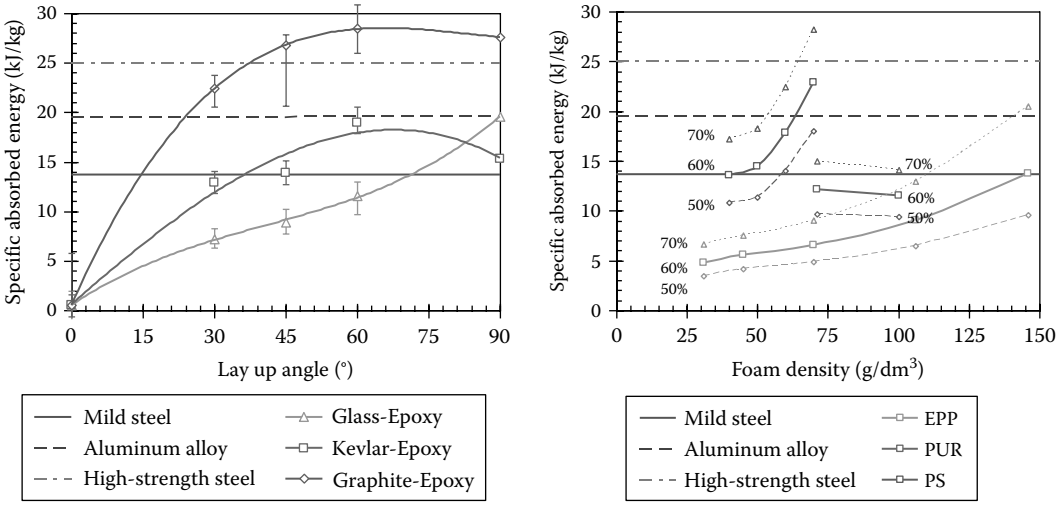


FIGURE 26.15 Typical specific absorbed energy for some structural car body materials.

As it is described in Chapter 43, in order to evaluate the safety level of the passengers, in a quantitative manner, biomechanical criteria are used. In particular, the actions on the head (through the *HIC*, Head Injury Criteria, index [99]), neck, and thorax, related to possible fatal consequences, and on the pelvis, legs, and femur and tibia bones, related to possible invalidating injuries, are considered [97]. The use of safety belts and air bags contributes to limit the motion of the passengers inside the passengers' compartment and, therefore, contributes to control the values reached by the biomechanical parameters, ensuring that they stay below the limit values indicated by the regulation [93].

Recently, a further regulation that accounts for possible impact of a vehicle against pedestrians has been developed [101]. Once again, some biomechanical parameters are used to verify that the level of possible injury suffered by the pedestrian is maintained below limit threshold [102]. The vehicle structure is involved with particular reference to the front bumper and the hood. For these two components, limits are prescribed relating to their aggressivity against the femur (front bumper) and the head (hood) [93].

26.7 FATIGUE ENDURANCE

Fatigue failures are determined by the application in service of loads of variable amplitude to the body structure [103,104].

The analysis of the fatigue strength of such a complex structure is quite difficult because it depends mainly on the following

- The real load history applied to the structure and its variability caused by the different driving conditions and styles
- The material fatigue strength and its natural dispersion
- The stress concentration effects that are induced by geometry details in the structure design and assembly joints
- The manufacturing process and the modification that it induces into the fatigue strength characteristics of the material, due, for example, to residual stresses

In order to have sufficient information about the real load histories that vehicles are submitted to during real-life riding, it is common to instrument some vehicle and to acquire signals of the applied loads and of the resulting strain fields when riding in typical riding conditions on typical roads. This procedure implies the determination of the vehicle's typical mission and utilization model [105]. Figure 26.16 shows an example of a resulting load spectrum that could be used for testing.

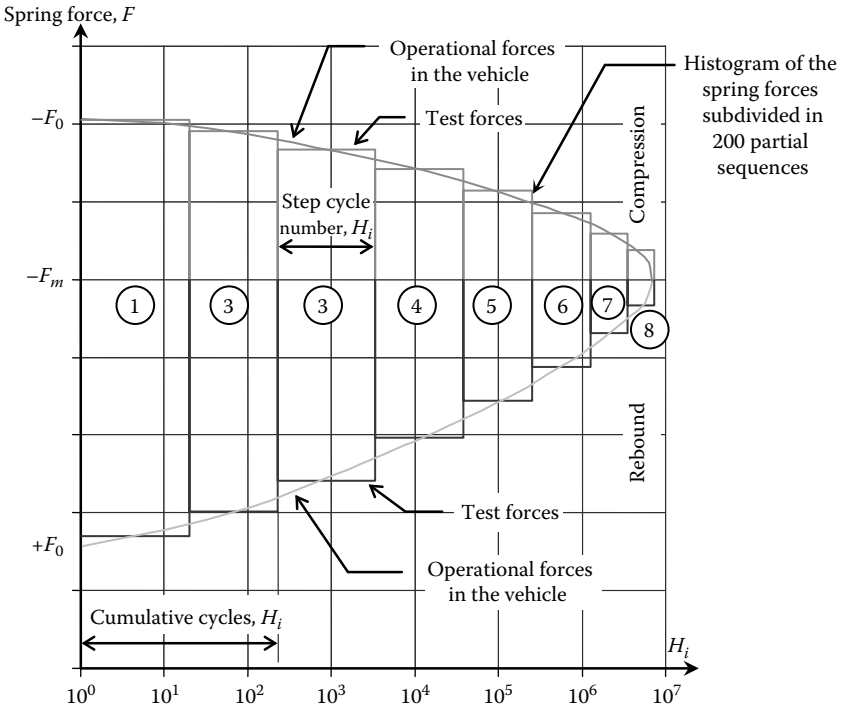


FIGURE 26.16 Load spectrum for fatigue test.

The acquired signals are processed in order to condense the information contained in the time histories and to get the so-called load cumulative, that is, the equivalent spectra of the stress mean and amplitude values. In the automotive industry, this signal processing is generally made by the rain flow algorithm [106–109], resulting in a three-dimensional matrix that accounts for stress mean values, stress oscillation amplitude values, and the number of times [106–110].

In recent years, the relevance of another factor on the fatigue behavior was pointed out: the loading frequency. It has become clear that load must be applied not simply to the "static" structure, as is traditionally done, but taking into account the vibration deformed shape that can be excited by the fatigue loads, due to their variability in time. Therefore, signal processing has been improved by the extraction of the coupled frequency information, resulting in a four-dimensional matrix [108].

This picture of real-life loading is used both for the design development through numerical simulations and for the laboratory experimental tests [104].

Road loading simulators [111] are nowadays in common use at the automotive industry testing laboratories due to the repeatability of the loading condition, ease of surveillance, and the rapidity with which the fatigue test can be completed, with respect to the tradition of road tests [111].

Fatigue performance of the materials has been analysed, with particular reference to the typical dispersion of the fatigue data, in order to have the so-called SNP (stress, number of cycles, probability) Wöhler curves [106–109]. In many cases, the companion curves, based on the strain (instead of stress) Wöhler-like curves, have been experimentally obtained, in order to take into account in a more appropriate way the plastic deformation content of each fatigue cycle [112,113].

Furthermore, the fatigue-testing procedure based on the thermoelastic and thermoplastic effects has been developed, as demonstrated by the correlation between the local temperature field and the fatigue failure localization [114].

As for all mechanical structures, the typical locations for fatigue failures in a car body are found where geometry determines stress concentration effect.

There are two structural details that are commonly of relevance in this perspective:

- The excessive sharpness of a sheet bent, where the geometry effect is generally coupled with material hardening due to the drawing plastic deformation
- The weld spot

In particular, this latter theme has been widely studied both from the theoretical and the experimental point of view.

The effect of stress concentration of the weld spot in the structure has been considered with the classical fatigue approach and with the fracture mechanic approach [115–120]. A good correlation between the experimental results and the finite element simulation results has been obtained by the application of the so-called hot-spot methodology [121], in order to identify a position for the evaluation of the actual stress entity, and the development of specialized finite elements [120].

The fatigue analysis has then been extended from the single weld spot to the behavior of the welded structural elements, where the interaction between the spot weld and the part geometry on one side and the entity of the weld pitch on the other side have relevant effects on the structure endurance.

In view of the adoption of the joining technique other than the spot welding, fatigue analysis has also been performed on adhesive joints [74,75] and on bolted joints [122]. These innovative joining methods have fatigue strength comparable to or better than spot welds, and, unexpectedly, adhesives show even better long-term behavior (Figure 26.17) [122,123].

The manufacturing process at least leaves the manufactured parts with important alteration with respect to the nominal conditions [124].

Deep drawing induces sheets reduction of the wall thickness of variable entity, somewhere very large, increment of the yielding strength and reduction of the metal ductility due to strain hardening, residual stresses of particular relevance when the elastic spring-back is constrained.

All these alterations of the real structure situation with respect to the nominal one lead to modification of the fatigue behavior of the considered structure that in a even more refined fatigue analysis should be accounted for [125].

Data on fatigue and fracture mechanics properties of the most commonly used materials for car body construction are listed in Table 26.4.

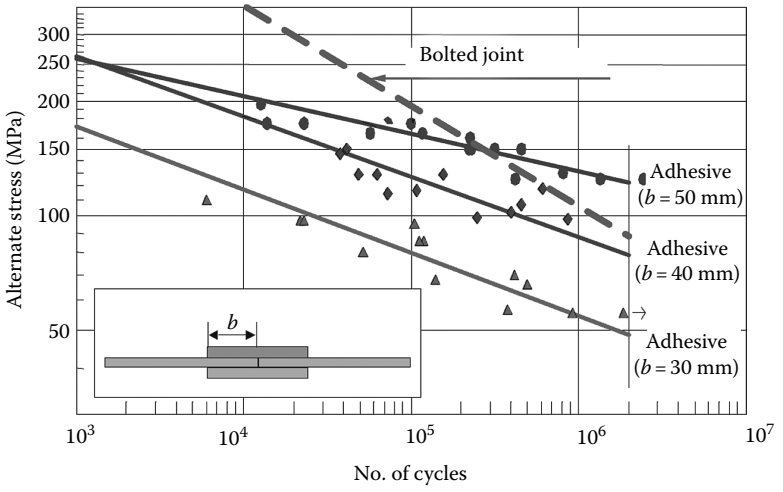


FIGURE 26.17 Comparison of the fatigue strength of some different joining techniques for car body structures. (From Lazzarin, P. et al., *Int. J. Fatigue*, 19, 401, 1997.)

TABLE 26.4
Fatigue and Fracture Mechanics Properties of the Most Commonly Used Materials for Car Body Construction

Reference	Material	E (GPa)	YS (MPa)	UTS (MPa)	A_f (%)	K_{1e} (MPa √m)	K_{1c} (MPa √m)	C	n	$\sigma_{D,-1}$ (MPa)
	DC01 EN10130		280	270–410	28			6.90E-12	3	
	DC03 EN10130		240	270–370	34			6.90E-12	3	
	DC04 EN10130		210	270–350	38			6.90E-12	3	
	DC05 EN10130		180	270–330	38			6.90E-12	3	
	DC06 EN10130		180	270–350	38			6.90E-12	3	
	S185 EN10027–1	207	185							
	S235 EN10027–1	207	235	325				2.43E-12	3.3	≥170
	S275 EN10027–1	207	275	360–510	24	110	77	1.91E-11	3	≥210
	S355 EN10027–1	207	355	450–610	22					≥280
[ULSAB-AVC]	DP500		300	500	30–34					
[ULSAB-AVC]	DP600		350	600	24–30					230–290
[ULSAB-AVC]	DP700		400	700	19–25					
[ULSAB-AVC]	DP800		500	800	14–20					310
[ULSAB-AVC]	TRIP 450/800		450	800	26–32					330–360
	TRIP 1000			1000						450
[ULSAB-AVC]	HSLA 350/450		350	450	23–27					
[ULSAB-AVC]	HSLA 490/600		490	600	21–26					
[45]	AA5083-O	71	138		43	69	49	5.27E-09	1.938	
[26]	AA6063-T5	69	145		27	37	26	1.73E-10	3.447	
[26]	A356-T60	72	213		40	24	18	2.56E-12	4.98	
	QE22A-T6 (Mg)	45	179		33	19	14	5.52E-09	3.21	

Sources: Rossetto, M., *Introduzione Alla fatica dei Materiali e dei Componenti Meccanici*, Levrotto & Bella, Torino, Piedmont, Italy, 2000; NASGRO Reference Manual: Fatigue crack growth computer program "NASGRO" version 3.0, National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, TX, Doc. No. JSC-22267B, 2000; Bauer, H., *Automotive Handbook*, Bosch, Stuttgart, Germany, 200; Finzi, A. and Valente, S., Potential applications of new HSS grade for wheels, *Proceedings of the Super-High Strength Steels*, Associazione Italiana di Metallurgia, Rome, Italy, Paper no 180, 2005; Yan, B., Fatigue behavior of advanced high strength steels for automotive applications, *Great Design in Steel Seminar*, 2003, http://www.autosteel.org

E, Young's modulus; YS, yield strength; UTS, ultimate tensile strength; A_f, elongation at break; K_{1e}, effective fracture toughness; K_{1c}, plane strain fracture toughness; C, Paris crack growth stress ratio; N, Paris exponent; $\sigma_{D,-1}$, fatigue limit, $R = -1$.

REFERENCES

1. Scibor-Rylski, A.J. and Skyes, D.M., *Road Vehicle Aerodynamic*, Pentech, London, U.K., 1984.
2. Dorgham, M.A., *Impact of Aerodynamic on Vehicle Design*, Interscience, Channel Islands, U.K., 1983.
3. Hucho, W.H. et al., *Aerodynamic of Road Vehicles: From Fluid Mechanics to Vehicle Engineering*, SAE, Warrendale, PA, 1998.
4. Dominy, R., Body design: Aerodynamics, in Happian-Smith, J. (ed.), *An Introduction to Modern Vehicle Design,* SAE International, Warrendale, PA, 2002, pp. 111–123.
5. Mori, K., Laser welding application of car body in Nissan, SAE paper 2003-01-2864, 2003.
6. Ebner, R., Laser welding of structural component, SAE paper 951857, 1995.
7. Staufer, D.H., Lichtbogen-laserschweissen von aluminium am beispiel des phaeton und des neuen AUDI A8, *Proceedings of the 7th German and 2nd European Automotive Conference Joining in Automotive Lightweight Design*, Bad Neuheim/Frankfurt, Germany, 2003, pp. 186–201.
8. Sünkel, R., Der weg zum spaceframe 2015—Nur ein materieller wettbewerb zwischen Stahl, aluminium, magnesium und titan? *Proceedings of the 14th European Lightweight Car Body Conference Low Cost Concepts versus Lightweight Innovations*, Bad Nauheim, Frankfurt, Germany, 2005, pp. 60–81.
9. Gobetto, E., Caroli, A., Rinero, E., and Naclerio, L., An hybrid architecture for light commercial vehicles featuring steel profiles space-frame and plastic panels, *Proceedings 14th European Lightweight Car Body Conference Low Cost Concepts versus Lightweight Innovations*, Bad Neuheim/Frankfurt, Germany, 2005, pp. 177–200.
10. Lengauer, G., Voestalpine-division profilform—Freedom of design for custom roll forming in vehicle construction, *Proceedings of the 14th European Lightweight Car Body Conference Low Cost Concepts versus Lightweight Innovations*, Bad Nauheim, Frankfurt, Germany, 2005, pp. 201–224.
11. Fredin, K. and Lassl, G., Vision lightweight design car body 2015, *Proceedings of the 14th European Lightweight Car Body Conference Low Cost Concepts versus Lightweight Innovations*, Bad Neuheim, Frankfurt, Germany, 2005, pp. 269–285.
12. Kreimeyer, M., Laser joined steel-aluminium hybrid structures, *Proceedings of the 6th European Car Body Conference Aluminium-steel-hybrid structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 126–143.
13. Fay, P.A. and Suthurst, G.D., Redesign of adhesively bonded box beam sections for improved impact performance, *International Journal of Adhesion and Adhesives*, 10(3), 128–138, 1990.
14. Szymberski, M., Adhesives for bonding plastics, SAE paper 2000-01-0421, 2000.
15. Fu, M. and Mallick, P.K., Performance of adhesive joints in an automotive composite structure, SAE paper 2000-01-1131, 2000.
16. Mayer, B., Pre-treatment and adhesive bonding of aluminium substrates, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 183–205.
17. Kunc, V., Erdman, D., and Klett, L., *Hybrid Joining in Automotive Applications*, ORNL/TM-119587, Oak Ridge, TN, 2001.
18. Schönfeld, R., Crash resistant structural bonding, *Proceedings of the Meeting on Structural Adhesives,* Vercelli, Piedmont, Italy, 2006, http://www.univer.polito.it/adesivi
19. Born, P., Low cure epoxy structural adhesives, *Proceedings of the Meeting on Structural Adhesives,* Vercelli, Piedmont, Italy, 2006, http://www.univer.polito.it/adesivi
20. Peroni, L. and Avalle, M., Experimental investigation of the energy absorption capability of bonded crash boxes, N. Jones and C.A. Brebbia (eds.), *Proceedings of the Ninth International Conference on Structures under Shock and Impact*, New Forest, U.K., 2006, pp. 445–454.
21. Pawlowski, J., *Vehicle Structures*, Cranfield University, Cranfield, U.K., 1986.
22. Fenton, J., *Handbook of Vehicle Design Analysis*, MPE Mechanical Engineering Publications, London, U.K., 1996.
23. Brown, J.C., Robertson, A.J., and Serpento, S.T., *Motor Vehicle Structures: Concepts and Fundamentals*, Butterworth Heinemann, Oxford, U.K., 2002.
24. Robertson, J., Chassis design and analysis, in Happian-Smith, J. (ed.), *An Introduction to Modern Vehicle Design,* SAE International, Warrendale, PA, 2002, pp. 125–155.
25. Zimnol, R. and Malek, Th., Weiterentwickelte kunststoffe fur den modernen karosseriebau: AuBenhaut, Verglasung und Karosseriestrukturen, *Proceedings of the 5th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 131–155.
26. Saito, M., Iwatsuki, S., Yasunaga, K., and Andoh, K., Development of aluminum body for the most efficient vehicle, *SAE of Japan Review*, 21, 511–516, 2000.

27. Wheeler, J., Crashworthiness of aluminum structured vehicles, *Proceedings of the 16th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Windsor, Ontario, Canada, Paper No. 98-Sl-W-20, 1998.
28. Walia, S., Gowland, S., Wakelin, P.G., Beckett, M., and Hemmings, J., The engineering of a body structure with hydroformed components, SAE paper 1999-01-3181, 1999.
29. Davies, G., *Materials for Automobile Bodies*, Elsevier, Oxford, U.K., 2003.
30. Vahl, M., Hein, P., and Bobbert, S., Hydroforming of sheet metal pairs for the production of hollow bodies, *Revue de Métallurgie*, 10, 1255–1263, 2000.
31. Lengauer, G., Leichtbau durch custom Rollforming im Fahrzeugbau,, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 380–407.
32. Le Flour, J.-C., The expected evolution of panels, profiles and pipes from the Renault perspective, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 10–31.
33. Castor, L., The challenge for aluminium in large series production, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 100–123.
34. Lahaye, Chr.T.W., Creating aluminium and steel opportunities for lightweight car body design, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 152–184.
35. Fredin, K., Future automobile materials from the Volvo car corporation perspective—Body applications, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 199–213.
36. Lozach, G., Lightweight materials in future automobile bodies of PSA Peugeot Citroën, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 279–313.
37. Avalle M. and Belingardi, G., Advanced materials for automotive applications, *Mobility & Vehicle Mechanics*, 30, 51–66, 2004.
38. Sörqvist, E.-L., Lightweight bumper system—Through high performance stainless steel and new forming methods, *Proceedings of the 14th European Lightweight Car Body Conference Low Cost Concepts versus Lightweight Innovations*, Bad Nauheim, Frankfurt, Germany, 2005, pp. 250–268.
39. Larsson, J.K., Design and manufacturing aspects of boron alloyed ultra high strength steel applications, *Proceedings of the 7th German and 2nd European Automotive Conference Joining in Automotive Lightweight Design*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 158–185.
40. Cazes, Chr. and Luciani, M., Innovative steel solutions for body-in-white and closures, *Proceedings of the 8th European Lightweight Car Body Design Conference Future Steel Car Bodies in Hybrid Design*, Bad Nauheim, Frankfurt, Germany, 2004, pp. 11–92.
41. Ikeda, S., A new advanced high strength steel, particularly developed for automotive seats, *Proceedings of the 8th European Lightweight Car Body Design Conference Future Steel Car Bodies in Hybrid Design*, Bad Nauheim, Frankfurt, Germany, 2004, pp. 213–227.
42. Takita, M. and Ohashi, H., Application of high-strength steel sheets for automobiles in Japan, *Revue de Métallurgie*, Paris, 10, 899–909, 2001.
43. Federici, C., Maggi, S., and Rigoni, S., The use of advanced high strength steel sheets in the automotive industry, *Proceedings of the Super-High Strength Steels*, Associazione Italiana di Metallurgia, Rome, Italy, Paper No. 11, 2005.
44. Bottema, J., Lahaye, C., Baartman, R., Zhuang, L., De Smet, P., and Schoepen, F., Recent developments in Aa6016 aluminum type body sheet product, SAE paper 981007, 1998.
45. van der Hoeven, J., Zhuang, L., Schepers, B., De Smet, P., and Baekelandt, J.P., A new 5xxx series alloy developed for automotive applications, SAE paper 2002-01-2128, 2002.
46. Schepers, B., De Smet, P., Baekelandt, J.P., van der Hoeven, J.A., and Zhuang, J., A new 5xxx series alloy developed for automotive applications, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 10–40.
47. Wieser, D., Bleche und Rohre aus naturharten aluminiumwerkstoffen fuer den Einsatz im Fahrwerksbereich,, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 41–60.
48. Bassi, C. and Timm, J., Last developments in the use of Al in car body construction, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 61–97.
49. Brünger, E., Maβgeschneiderte Al- Bleche für Türen und Klappen, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 98–125.

50. Hirsch, J., Aktuelle und zukuenftige Einsatzpotentiale von Aluminium und Magnesium, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 124–151.
51. Blanchard, P.J. et al., The application of magnesium die casting to vehicle closures, SAE paper 2005-01-0338, 2005.
52. Feraboli, P. and Masini, A., Development of carbon/epoxy structural components for a high performance vehicle, *Composites—Part B Engineering*, 35, 323–330, 2004.
53. Clemente, R., Miravete, A., Larrodé, E., and Castejón, L., 3D composite sandwich structures applied to car manufacturing, SAE paper 980071, 1998.
54. Schmachtenberg, E. and Kleba, I., Status and future development concerning the application on fibre reinforced components for car body structure from the scientific point of view, *5th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 10–54.
55. Kopp, J. and Aitken, C., CFRP bodywork—Volume-oriented cost-efficiency and class A quality, *Proceedings of the 6th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 10–46.
56. Diboine, A., Plastic parts in Renault's production cars—Past, present and future, *Proceedings of the 6th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 128–173.
57. Cramer, D., Design and manufacture of a cost-effective advanced-composite automotive body structure, *5th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 156–179.
58. Grunewald, U., Advanced sandwich materials (Multifunctional Materials)—Opportunity for future lightweight vehicles, *Proceedings of the 8th European Lightweight Car Body Design Conference Future Steel Car Bodies in Hybrid Design*, Bad Nauheim, Frankfurt, Germany, 2004, pp. 336–355.
59. Renault, Th., Application of lightweight composite sandwich structures to the floorpan of Microcar MC1, *Proceedings of the 6th European Car Body Conference Vision Plastic Car Body 2010*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 191–206.
60. Fox-Rubin, J., Fiber-reinforced plastics design and manufacturing for structural parts, *Proceedings of the 1st European Strategy Conference Vision of Materials and Processes*, Berlin, Germany, 2002, pp. 400–414.
61. Barnes, T.A. and Pashby, I.R., Joining techniques for aluminium spaceframes used in automobiles Part I—Solid and liquid phase welding, *Journal of Materials and Processing Technology*, 99, 62–71, 2000.
62. Barnes, T.A. and Pashby, I.R., Joining techniques for aluminium spaceframes used in automobiles Part II—Adhesive bonding and mechanical fasteners, *Journal of Materials and Processing Technology*, 99, 72–79, 2000.
63. Kochan, A., Lotus: Aluminium extrusions and adhesives, *Assembly Automation*, 16(4), 19–21, 1996.
64. Belingardi, G., Recent advances in joining technology for car body application, *Proceedings of the International Congress Motor Vehicles &Motors*, Kragujevac, Serbia, Paper no MVM2006IL01, 2006.
65. Kulak, G.L., Fisher, J.W., and Struik, J.H., *Guide to Design Criteria for Bolted and Riveted Joints*, Wiley, New York, 1987.
66. Ogawa, T., Toyonaga, S., Tanigawa, O., and Kato, T., Development of sleeve clinching method and making practicable, SAE paper 970372, 1997.
67. Mortimer, J., Jaguar uses X350 car to pioneer use of self-piercing rivets, *Industrial Robot: An International Journal*, 28(3), 192–198, 2001.
68. Mortimer, J., Jaguar uses castings, extrusions to reduce parts count in new sports car, *Assembly Automation*, 26(2), 115–120, 2006.
69. Hunter, J.A. et al., Design and production of adhesively bonded aluminium automotive structures, *Proceedings of the 31st Conference on ISATA*, Dusseldorf, Germany, 1998.
70. Beevers A. et al., Analysis of stiffness of adhesive joints in car body, *Journal of Material Processing and Technology*, 118, 96–101, 2001.
71. Mortensen, F. and Thomsen, O.T., Analysis of adhesive bonded joints: A unified approach, *Composite Science and Technology*, 62, 1011–1031, 2002.
72. Belingardi, G. and Chiandussi, G., Stress flow in thin-walled box beams obtained by adhesive bonding joining technology, *International Journal of Adhesion and Adhesives*, 24, 423–439, 2004.
73. Krenk, S., Jonsson, J., and Hansen, L.P., Fatigue analysis and testing of adhesive joints, *Engineering Fracture Mechanics*, 53, 859–872, 1996.
74. Belingardi, G., Avalle, M., Fizzotti, R., and Tarditi, A., Experimental results on the axial crushing behaviour of hat-section thin-walled beam assembled by adhesive bonding, *Proceedings of the 2001 ATA Florence Conference*, ATA paper 01A1043, 2001.

75. De Goeij, W.C., van Tooren, M.J.L., and Beukers, A., Composite adhesive joints under cycling loading, *Materials and Design*, 20, 213–221, 1999.
76. Kaya, A., Tekelioğlu, M.S., and Findik, F., Effects of various parameters on dynamic characteristics in adhesively bonded joints, *Materials Letters*, 58, 3451–3456, 2004.
77. Krois, M., Dilger, K., Böhm, S., and Koch, S., Use of compliant adhesive layers in direct glazing of road vehicles for improved sound attenuation, *International Journal of Adhesion & Adhesives*, 23, 413–425, 2003.
78. Mortimer, J., Adhesive bonding of car body parts by industrial robot, *Industrial Robot: An International Journal*, 31(5), 423–428, 2004.
79. Ahlers-Hestermann, G., Fortschritte beim mechanischen Fügen von Aluminium und Stahl/Aluminium, *Proceedings of the 6th European Car Body Conference Aluminium-Steel-Hybrid Structures*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 159–182.
80. Backe, P., Stanznieten mit thermischer Unterstuetzung zum Fuegen Magnesiumblechen, *Proceedings of the 7th German and 2nd European Automotive Conference Joining in Automotive Lightweight Design*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 80–100.
81. Mauermann, R. and Dietrich, S., Entwicklungsfortschritte beim matrizen—Losen Clinchen und Nietclinchen, *Proceedings of the 7th German and 2nd European Automotive Conference Joining in Automotive Lightweight Design*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 50–79.
82. Sowa, C., Hybrid—Fügetechnik—Kleben/Stanznieten am Beispiel der Rohbautür des Jaguar XJ (X 350), *Proceedings of the 7th German and 2nd European Automotive Conference Joining in Automotive Lightweight Design*, Bad Nauheim, Frankfurt, Germany, 2003, pp. 10–49.
83. Brughmans M. et al., The application of FEM-EMA correlation and validation techniques on body in white, *IMechE '94*, Paper C487/003/94, 1994.
84. Garro, A. and Belingardi, G., Tecniche di discretizzazione nel progetto dell'autoveicolo, *Il progettista industriale*, 3, 74–85, 1985.
85. Kitahara, T., Terada, I., and Watanabe, T., Study on effective application of soundproofing materials through low noise prototype car development, *International Journal of Vehicle Design*, 5(4), 490–504, 1984.
86. Genta, G., *Vibration of Structures and Machines*, 3rd edn., Springer, New York, 1998.
87. Garro, A., *Progettazione Strutturale del Motore*, Levrotto & Bella, Torino, Piedmont, Italy, 1992.
88. Yamanouchi, K., Yamashita, K., Mamiya, N., Yamazaki, T., and Yamazaki, I., Application of prediction noise and vibration analysis to the development of a new-generation lightweight 3-Liter V6 Nissan Engine, SAE paper 940993, 1994.
89. Lee, D.-H., Hwang, W.-S., and Kim, C.-M., Design sensitivity analysis and optimization of an engine mount system using an FRF-based substructuring method, *Journal of Sound and Vibration*, 255(2), 383–397, 2002.
90. Ewins, D.J., *Modal Testing: Theory, Practice and Application*, Baldock, Hertfordshire, U.K., 2000.
91. Nakanishi, Y. and Gerges, S.N.Y., Acoustic modal analysis for vehicle cabin, SAE paper 952246, 1995.
92. Maurin, M., A characterization of the acoustic emissions of moving vehicles, *Recherche Transports Sécurité*, 1998(59), 43–57, 1998.
93. Prasad, P. and Belwafa, J.E., *Vehicle Crashworthiness and Occupant Protection,* AISI, Southfield, MI, 2004.
94. Belingardi, G., Some recent development and perspective in car passive safety, *Mobility & Vehicle Mechanics*, 28, 87–105, 2002.
95. European New Car Assessment Program (EuroNCAP) Frontal Impact Testing Protocol v4.1, 2004. http://www.euroncap.com/downloads/test_procedures/area_3/event_2/EuroNCAP%20Frontal%20Protocol%204.1.pdf
96. European New Car Assessment Program Protocol Changes and Additions—March 2004. http://www.euroncap.com/downloads/test_procedures/area_3/event_7/Protocol%20Changes%2025th%20March%2004.pdf
97. ULSAB-AVC Technical Transfer Dispatch #6, CAE Analysis for Crashworthiness, 1999.
98. Sugimoto, T., Kadotani, Y., and Ohmura, S., The offset crash test—A comparative analysis of test methods, *Proceedings of the 16th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Windsor, Ontario, Canada, Paper no 98-Sl-O-08, 1998.
99. DeWeese, R.L. and Moorcroft, D.M., *Evaluation of a Head Injury Criteria Component Test Device*, U.S. Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Document DOT/FAA/AM-04/18, 2004.
100. Seiffert, U. and Wech, L., *Automotive Safety Handbook*, SAE International, Warrendale, PA, 2003.

101. UNECE World Forum for Harmonization of Vehicle Regulations (WP.29), Working Party on Passive Safety (GRSP): Proposal for a global technical regulation on uniform provisions concerning the approval of vehicles with regard to their construction in order to improve the protection and mitigate the severity of injuries to pedestrians and other vulnerable road users in the event of a collision, Document ECE/TRANS/WP.29/GRSP/2006/2 http://www.unece.org/trans/doc/2006/wp29grsp/ECE-TRANS-WP29-GRSP-2006-02e.pdf
102. European New Car Assessment Program (EuroNCAP) Pedestrian Testing Protocol v4.1, 2004 http://www.euroncap.com/downloads/test_procedures/area_3/event_2/EuroNCAP%20Pedestrian%20Protocol%204.1.pdf
103. Wetzel, R.M., Fatigue under complex loading: analysis and experiments, SAE International, Warrendale, PA, 1977.
104. Kondo, Y. et al., Prediction methods of rough-road-load applied to vehicle body, SAE paper 930665, 1993.
105. Hay, N.C. and Dodds, C.J., Vehicle load histories: The duality of vibration and fatigue spectra, SAE paper 871938, 1987.
106. Fuchs, H.O. and Stephens, R.I., *Metal Fatigue in Engineering*, New York, 1980.
107. Collins, J.A., *Failure of Materials in Mechanical Design: Analysis, Prediction, Prevention*, Wiley, New York, 1993.
108. Rossetto, M., *Introduzione Alla fatica dei Materiali e dei Componenti Meccanici*, Levrotto & Bella, Torino, Piedmont, Italy, 2000.
109. Rice, R.C., *Fatigue Design Handbook*, R.C. Rice (ed.), SAE International, Warrendale, PA, 1988.
110. Socie, D. and Park, K., Analytical description of service loading suitable for fatigue analysis, SAE paper 971535, 1997.
111. Davis, A.G.W. and Asher, G., Control strategy for a multi-roll chassis dynamometer in road load simulation, SAE paper 950514, 1995.
112. NASGRO Reference Manual: Fatigue crack growth computer program "NASGRO" version 3.0, National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, TX, Doc. No. JSC-22267B, 2000.
113. Durán, J.R. and Pinho de Castro, J.T., Prediction of fatigue crack propagation under complex loading using εN parameters, SAE paper 2001-01-4064.
114. La Rosa, G., and Risitano, A., Thermographic methodology for rapid determination of the fatigue limit of materials and mechanical components, *International Journal of Fatigue*, 22(1), 65–73, 2000.
115. El-Sayed, M.E.M., Stawiarski, T., and Frutiger, R., Fatigue analysis of spot-welded joints under variable amplitude load history, *Engineering Fracture Mechanics*, 55, 363–369, 1996.
116. Zhang, S., Stress intensity at spot welds, *International Journal of Fracture*, 88, 281–294, 1997.
117. Blarasin, A. et al., Design fatigue data for joining steel sheets in the automotive industry, SAE paper 900742, 1990.
118. Lin S.H. et al., Spot weld failure loads under combined mode loading conditions, SAE paper 2001-01-428, 2001.
119. Zhang, S., Recent development in analysis and testing of spot welds, SAE paper 2001-01-0432, 2001.
120. Salvini, P., Vivio, F., and Vullo, V., A spot weld finite element for structural modelling, *International Journal of Fatigue*, 22, 645–656, 2000.
121. Radaj, D., Review of fatigue strength assessment of non-welded and welded structures based on local parameters, *International Journal of Fatigue*, 18(3), 153–170, 1996.
122. Lazzarin, P., Milani, V., and Quaresimin, M., Scatter band summarising the fatigue strength of aluminium alloy bolted joints, *International Journal of Fatigue*, 19, 401–407, 1997.
123. Blarasin, A., Caffa, G., Mennicucci, G., and Gastaldi, G., Proprietà meccaniche e a fatica di giunzioni per scocca vettura realizzate con tecnologie innovative, *Internal Report* CRF AA55/1, 1986.
124. Chiandussi, G. et al., Analysis of the effects of the manufacturing process in the mechanical behaviour of an automotive bumper, *Proceedings of the SheMet 2001—9th International Conference on Sheet Metal*, Leuven, Belgium, 2001, pp. 43–50.
125. Murakami, Y. and Nagata, J., Effects of small defects on fatigue properties of materials, mechanism of ultralong life fatigue and applications to fatigue design of car components, SAE paper 2002-01-0575, 2002.
126. Bauer, H., *Automotive Handbook—3rd edn*., H. Bauer (ed.), Bosch, Stuttgart, Germany, 200.
127. Finzi, A. and Valente, S., Potential applications of new HSS grade for wheels, *Proceedings of the Super-High Strength Steels*, Associazione Italiana di Metallurgia, Rome, Italy, Paper no 180, 2005.
128. Yan, B., Fatigue behavior of advanced high strength steels for automotive applications, *Great Design in Steel Seminar*, 2003, http://www.autosteel.org

Part V

Vehicle Dynamics and Active Safety

27 Basics of Longitudinal and Lateral Vehicle Dynamics

Manfred Ploechl, Peter Lugner, and Johannes Edelmann

CONTENTS

27.1 Introduction 971
27.2 Longitudinal Vehicle Behavior 972
27.2.1 Application and Vehicle Model Features 972
27.2.2 Vehicle and Drive Train Model 972
27.2.3 Driving Resistances 978
27.2.3.1 Rolling Resistance 978
27.2.3.2 Aerodynamic Phenomena and Drag 979
27.2.4 Vehicle Performance 982
27.2.4.1 Driving Performance 982
27.2.4.2 Braking 985
27.2.5 Reaction, Braking Procedure, and Stopping 988
27.2.6 State-Space Description, Inverse Dynamics, and Basics for Longitudinal Control 989
27.3 Lateral Vehicle Behavior 991
27.3.1 Application and Vehicle Model Features 991
27.3.2 Linear Two-Wheel Vehicle Model 992
27.3.2.1 Equations of Motion 992
27.3.2.2 Steady-State Cornering 995
27.3.2.3 Stability and Nonsteady-State Vehicle Behavior 1001
27.3.3 Nonlinear Two-Wheel Vehicle Model 1007
27.4 Subsystems and Vehicle–Trailer Combination 1013
27.4.1 Braking of a Wheel until Lock-Up 1013
27.4.2 Wheel Shimmy 1016
27.4.3 Vehicle–Trailer Combination 1018
References 1020

27.1 INTRODUCTION

For a basic understanding of the low-frequency behavior of an automobile, it is essential to model its corresponding characteristics in a simplified, mainly linearized way. As a consequence, analytical solutions or easy to interpret results can be used to analyze the vehicle behavior for a large area of driving and testing conditions, for example [1–11].

With these simplified vehicle models, it is possible to consider the longitudinal and lateral behavior of the vehicle for a large range of problems separately. Then, characteristic phenomena, for example, the design of the braking system or under- and oversteer behavior, can be understood without much effort.

Another field of application for linearized models is the area of vehicle dynamics control systems, since with these models feedback control gains and state observers can be derived and designed rather easily; for example, they constitute the theoretical basis of the electronic stability program (ESP) or of similar control systems.

Establishing an adequate vehicle model is a compromise of what is to be investigated and what restrictions and/or simplifications should determine the details of the vehicle model. Consequently, at the beginning of the sections on modeling, a list of possible applications and model assumptions not only give an overview of the characteristics of the vehicle model but also define its range of applications.

As most handling features can be shown with a passenger car (more easily), this type of vehicle is presented in the following.

27.2 LONGITUDINAL VEHICLE BEHAVIOR

27.2.1 Application and Vehicle Model Features

The "longitudinal" behavior of the vehicle is characterized by equal tire loads (tire normal forces) at the left and right side of the car or small differences that can be neglected. So driving along a curved trajectory can be analyzed assuming that the lateral accelerations or other conditions causing changes in left to right load transfer are small.

Consequently, the following objectives can be investigated:

1. Driving, starting, and braking on a road of arbitrary grade
2. Effects of different drive train configurations like front-wheel drive, rear-wheel drive, four-wheel drive
3. Basic brake system design
4. Brake systems with control system, for example, antilock braking system (ABS)
5. Interaction of driver–vehicle–road (traffic)
6. Different road surface properties including μ-split conditions
7. Engine performance, effects of (automatic) transmission, clutch, converter, and other drive train components
8. Fuel consumption, efficient driving

Some of these tasks not only require a vehicle model but also completions by models of the drive train and its components, the engine characteristics, and the longitudinal force transfer behavior of the tires.

With restriction to a plane vehicle model and an overall description of the drive train (including a simple engine and tire behavior description), the main objectives of points 1–3 and 6 can be solved. For point 5, the application of a driver model is necessary, see Chapter 38, and for point 4, nonlinear tire characteristics need to be included. A detailed model of the engine characteristics and its fuel consumption is necessary for the investigation of 7 and 8. With a focus on the overall vehicle behavior, the aforementioned tasks 1–3 and 6 are addressed here, while point 4, in particular, is considered in Chapter 30 of this handbook. For points 7 and 8, please refer to [7] and [12], respectively.

27.2.2 Vehicle and Drive Train Model

The vehicle is assumed to be represented by a plane rigid body model, where the relative motions of the wheels with respect to the car body, except their spins, are neglected (see Figure 27.1, e.g., [2,5,6]). The tire–road contact is described either by pure rolling and/or by its limitation with a maximum longitudinal contact force coefficient $\mu_{max} = (F_x/F_z)_{max}$, which is (to simplify matters) the same as μ_s for the locked or sliding wheel. For most of the basic longitudinal investigations, this tire force transfer model assumption is an efficient and reasonable approximation, because the longitudinal tire slip is small, when the wheel does not slide.

For establishing the equations of motion, the vehicle running on a grade (angle ϑ) with velocity v_x and acceleration a_x is considered.

At the center of gravity *CG* of the whole vehicle (mass *m*), the aerodynamic lift W_Z, drag force W_L, and lift moments M_Y act (see Section 27.2.3.2) besides weight *mg*. The tire forces F_{xi} and F_{zi} (i=*F*, *R*)

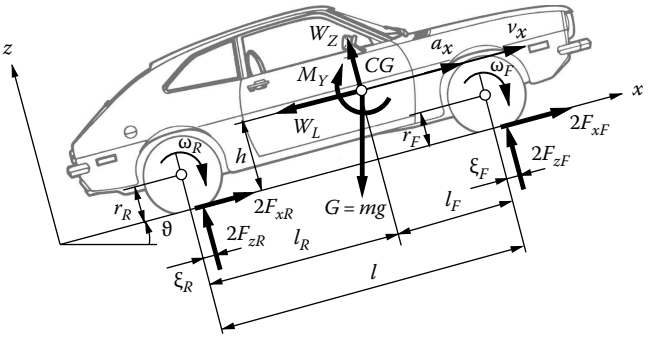


FIGURE 27.1 Plane longitudinal vehicle model.

corresponding to each wheel are assumed to be equal for the right- and left-hand side. The longitudinal shift ξ_i of F_{zi} (i=F, R) is the consequence of the rolling resistance of the tire resulting in the rolling resistance coefficient (see also Section 27.2.3.1)

$$f_{Ri} = \frac{\xi_i}{r_i} \tag{27.1}$$

which is assumed to be equal for all tires and named f_R.

Newton's law and Euler's equation with respect to CG constitute (for $G = mg$)

$$ma_x = 2F_{xF} + 2F_{xR} - W_L - G\sin\vartheta \tag{27.2}$$

$$0 = 2F_{zF} + 2F_{zR} + W_Z - G\cos\vartheta \tag{27.3}$$

$$2I_F\dot{\omega}_F + 2I_R\dot{\omega}_R = 2F_{zR}(l_R - \xi_R) - 2F_{zF}(l_F + \xi_F) - 2(F_{xF} + F_{xR})h + M_Y \tag{27.4}$$

with the moments of inertia I_F, I_R of the front and rear wheels, which rotate with the angular velocities ω_F, ω_R.

To get the drive-train and/or brake system involved, the corresponding drive and/or braking moments, MDi and/or MBi, are presented in Figure 27.2, acting at one rear wheel. The forces X_R, Z_R result from the force transfer between the car body and the center of the wheel. Aerodynamic effects on the wheels are not considered separately, but assumed to be taken into account by W_L, W_Z, M_Y in (27.2) through (27.4) (see Section 27.2.3.2). The corresponding Euler's equation w.r.t. CG_R is

$$I_R\dot{\omega}_R = M_{DR} - M_{BR} - M_{FR} - F_{zR}\xi_R - F_{xR}r_R \tag{27.5}$$

including the friction moments M_{FR} of the wheel bearing. To simplify the kinematics, as indicated before, and further assuming the same tires on the front and rear axle, a rolling condition of the wheel is established by

$$\begin{aligned} r\omega &= v_x \\ r\dot{\omega} &= a_x \end{aligned} \tag{27.6}$$

$$\begin{aligned} \text{with } r &= r_F = r_R \\ \omega &= \omega_F = \omega_R \end{aligned} \tag{27.7}$$

And the simplification that the effective rolling r is equal to the loaded radii $r_F = r_R$.

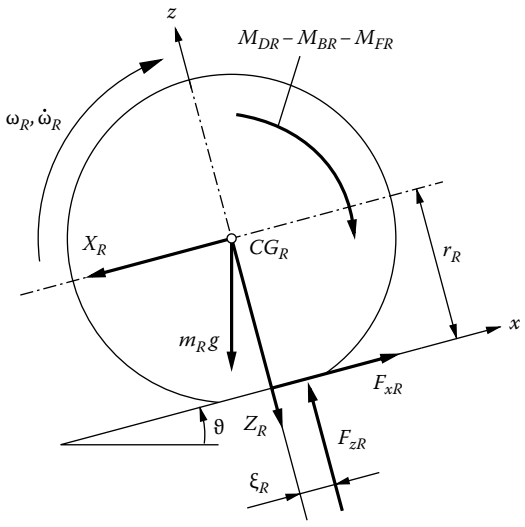


FIGURE 27.2 Forces and moments at the rear wheel.

Now, the longitudinal tire force F_{xR} in (27.5) can be expressed by using (27.1), and (27.7):

$$F_{xR} = \frac{1}{r}(M_{DR} - M_{BR} - M_{FR}) - \frac{I_R \dot{\omega}}{r} - f_R F_{zR} \tag{27.8}$$

With (27.8) and a corresponding equation for the front wheel, (27.2) reads

$$ma_x + 2I_R \frac{\dot{\omega}}{r} + 2I_F \frac{\dot{\omega}}{r} = \frac{1}{r}\left[(2M_{DR} + 2M_{DF}) - (2M_{BR} + 2M_{BF})\right]$$
$$-2\left[\left(\frac{M_{FR}}{r} + \frac{M_{FF}}{r}\right) + (f_R F_{zR} + f_R F_{zF})\right] - W_L - G\sin\vartheta$$

or by incorporation of (27.6) and (27.3)

$$\left(m + \frac{2I_R}{r^2} + \frac{2I_F}{r^2}\right) a_x = \frac{M_D}{r} - \frac{M_B}{r} - W_{total} \tag{27.9}$$

with the substitutes

$$W_{total} = \frac{M_F}{r} + W_R + W_L + W_G$$

$$W_R = f_R(2F_{zR} + 2F_{zF}) \overset{(27.3)}{=} f_R(G\cos\vartheta - W_Z)$$

$$W_G = G\sin\vartheta \tag{27.10}$$

$$M_F = 2(M_{FR} + M_{FF})$$

$$M_D = 2M_{DR} + 2M_{DF}$$

$$M_B = 2M_{BR} + 2M_{BF}$$

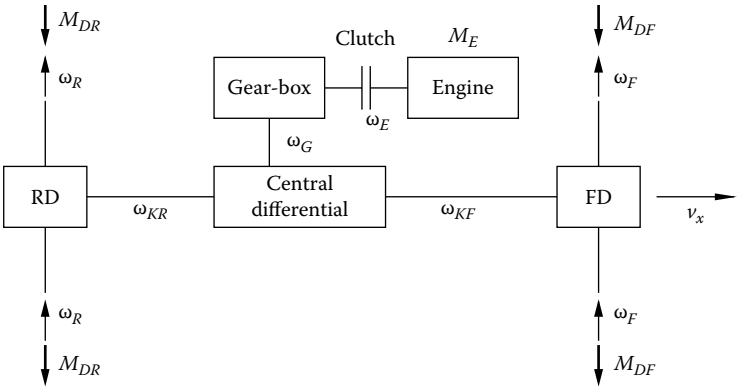


FIGURE 27.3 Schema of a four-wheel drive with manual transmission (RD, rear differential; FD, front differential).

Equation 27.9 reveals that the acceleration of the vehicle is determined by the sum of the driving torques M_D at the wheels, the braking torques M_B, and the whole driving resistance force W_{total} (see Section 27.2.3), comprising influences of the rolling resistance W_R, the grade by W_G, the aerodynamic drag W_L, and smaller effects by friction moments in the wheel bearings that are usually neglected.

The drive train is modeled in such a way that the different drive concepts—front-wheel drive, rear-wheel drive, and four-wheel drive—can be considered (Figure 27.3). In this figure, the wheels are omitted since the dynamics are already taken into account by (27.5). The same wheel-spin velocities ω_F, ω_R and driving torques M_{DF}, M_{DR} for both wheels of an axle hold for the presumed straight-line motion and symmetric conditions left–right. (For open differentials and, e.g., cornering the wheel spins, ω_i will be different for each wheel, while for locked differentials the drive moments will be different for each wheel.)

Considering the power–energy equation $\dot{E}_{kin} = \sum P$ and taking into account that the derivative of the translational kinetic energy is balanced by the corresponding power of the longitudinal external forces acting on the considered system, the derivative of the rotational kinetic energy

$$E_{rot} = \Theta_E \frac{\omega_E^2}{2} + I_C \frac{\omega_G^2}{2} + I_{DR} \frac{\omega_{KR}^2}{2} + I_{DF} \frac{\omega_{KF}^2}{2} \tag{27.11}$$

has to be only considered (Figure 27.3). Θ_E is the substitutive moments of inertia of the engine and all parts rotating with the angular velocity ω_E of the engine; I_C, the moments of inertia of parts of the transmission and central differential rotating with ω_G; and I_{DR}, I_{DF}, the moments of inertia of the rear and front differential and parts rotating with ω_{KR}, ω_{KF}, respectively. The moments of inertia of the drive shafts from differential to wheels are assumed to be included in I_R, I_F. Thereby, the power–energy equation results in

$$\frac{dE_{rot}}{dt} = \eta M_E(\omega_E)\omega_E - 2M_{DF}\omega_F - 2M_{DR}\omega_R \tag{27.12}$$

The input power $M_E\omega_E$ of the engine torques $M_E(\omega_E)$ is reduced by applying a coefficient of efficiency $\eta < 1$, presenting the transfer losses in the system.

By means of the central differential, a distribution of the driving torques M_D to the cardan shafts can be accomplished:

$$2M_{DF} = M_D \nu_F, \quad 2M_{DR} = M_D \nu_R$$
$$\nu_F + \nu_R = 1 \tag{27.13}$$

$$\begin{aligned} \text{for rear-wheel drive:} \quad & \nu_R = 1 \\ \text{for front-wheel drive:} \quad & \nu_F = 1 \\ \text{for 4-WD with equal distribution:} \quad & \nu_R = \nu_F = 0.5 \end{aligned} \tag{27.14}$$

By introducing transmission ratios N_{Gn} of the gears and N_D of the differentials, the following kinematic relations are obtained:

$$\begin{aligned} \omega_E &= \omega_G N_{Gn} \\ \omega_G &= \nu_R \omega_{KR} + \nu_F \omega_{KF} \\ \omega_{KR} &= N_D \omega_R, \quad \omega_{KF} = N_D \omega_F \end{aligned} \tag{27.15}$$

The transmission ratio N_{Gn} is determined by the chosen gear.

With the simplifications (27.6) and (27.7) and the relations in (27.13) and (27.15), Equation 27.12 provides the driving torques

$$\begin{aligned} M_D &= 2M_{DF} + 2M_{DR} \\ &= \eta M_E(\omega_E) \cdot N - r \left[\frac{\Theta_E N^2}{r^2} + \frac{(I_C + I_{DR} + I_{DF})}{r^2} N_{Gn}^2 \right] a_x \end{aligned} \tag{27.16}$$

with

$$N = N_{Gn} N_D$$

In (27.16), it is assumed that N_{Gn} is not a function of time (as it may happen with continuous gear transmissions), and that there is no instantaneous change due to an action of the clutch or a converter.

An essential assumption for this kind of drive train model is (27.6), (27.7), where no slip–force relation at the tire is considered. An extension of this drive train model, applicable also for cornering, is given in [9].

Finally, inserting (27.16) into (27.9) yields the relation between engine torques input and resulting acceleration of the vehicle

$$(m + m_r) a_x = \frac{\eta M_E \omega_E \cdot N}{r} - \frac{M_B}{r} - W_{total} \tag{27.17}$$

with the equivalent mass of the rotating components

$$m_r = \frac{1}{r^2} \left[\Theta_E N^2 + (I_C + I_{RD} + I_{FD}) N_{Gn}^2 + (2I_R + 2I_F) \right] \tag{27.18}$$

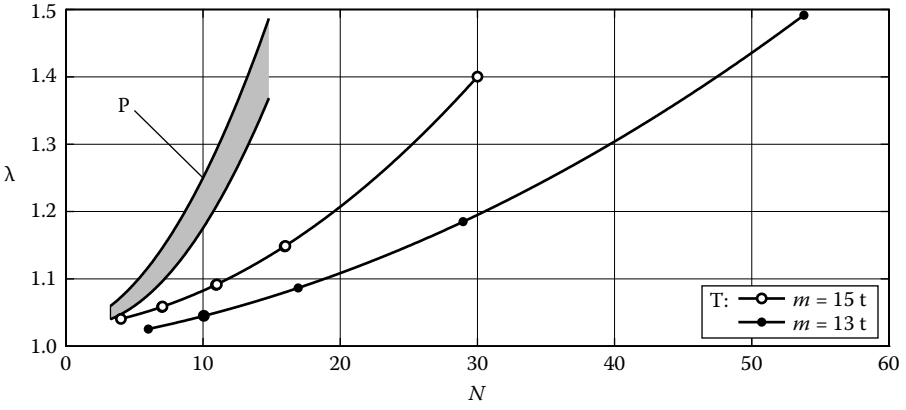


FIGURE 27.4 Typical mass factor λ as a function of the overall transmission ratio N for passenger cars (P) and trucks (T).

Due to the $N = N_{Gn}N_D$, the mass factor

$$\lambda = \frac{m + m_r}{m} \tag{27.19}$$

depends to a great extent on the applied gear ratio N_{Gn} (see Figure 27.4, [5]).

For the determination of the longitudinal force coefficient of wheel i,

$$\mu_i = \frac{F_{xi}}{F_{zi}} \tag{27.20}$$

the normal tire forces F_{zi} have to be derived from (27.3) and (27.4).

$$0 = 2F_{zF} + 2F_{zR} + W_Z - G\cos\vartheta$$

$$2I_F\dot{\omega}_F + 2I_R\dot{\omega}_R + ma_x h = -2F_{zF}l_F + 2F_{zR}l_R + (-W_L - G\sin\vartheta)h + M_Y - 2(\xi_F F_{zF} + \xi_R F_{zR})$$

by using $(2F_{xF} + 2F_{xR})$ from (27.2). With $\xi_F = \xi_R = \xi$ and (27.6), (27.7), the aforementioned two equations result in

$$2F_{zF} + 2F_{zR} = G\cos\vartheta - W_Z \tag{27.21}$$

$$-2l_F F_{zF} + 2l_R F_{zR} = \left[\left(\frac{2I_F}{r} + \frac{2I_R}{r}\right)a_x + (ma_x + W_L + G\sin\vartheta)h + M_Y + \xi(G\cos\vartheta - W_Z)\right] \tag{27.22}$$

and with a_x from (27.17), the normal tire forces can be determined explicitly. Often, only the main components of the right-hand sides are considered: $M_Y \approx 0$, $W_Z \approx 0$, and by neglecting

the small terms $(2I_F + 2I_R)/r \ll mh$ as well as $\xi(G\cos\vartheta) \ll (ma_x + W_L + G\sin\vartheta)h$, the normal tire forces follow from

$$\frac{F_{zF}}{G} = \frac{l_R}{l}\cos\vartheta - a^* \frac{h}{l} \tag{27.23}$$

$$\frac{F_{zR}}{G} = \frac{l_F}{l}\cos\vartheta + a^* \frac{h}{l} \tag{27.24}$$

$$\text{with } a^* = \left(\frac{a_x}{g} + \sin\vartheta + \frac{W_L}{G} \right) \tag{27.25}$$

where, moreover, for common grades $|q| \leq 20\%$ (respectively, $|\tan\vartheta| \leq 0.2$), the approximation $\cos\vartheta \approx 1$ is appropriate (error $\leq 2\%$). The tire load transfer front–rear, with respect to acceleration a_x and grade ϑ, can now be easily evaluated.

27.2.3 Driving Resistances

Sometimes expressions like "acceleration resistance" and "resistance due to a grade"—which can by definition be positive or negative—can be found in literature, but need not be considered separately, since these phenomena are already included in the formulations of Section 27.2.2, for example, by "ma_x" and "$G \sin\vartheta$." The friction moments M_F caused by the wheel bearing can generally be neglected compared, for example, to the rolling resistance [7].

27.2.3.1 Rolling Resistance

The rolling resistance originates from the energy dissipation of the deformation of the rolling tire [7]. Thereby, the pressure distribution in the tire contact patch has larger values in the direction of rolling (see Figure 27.5), which results in a normal tire force acting in the front of the virtual tire contact point A. This effect is already taken into account in Figures 27.1 and 27.2 by the introduction of ξ.

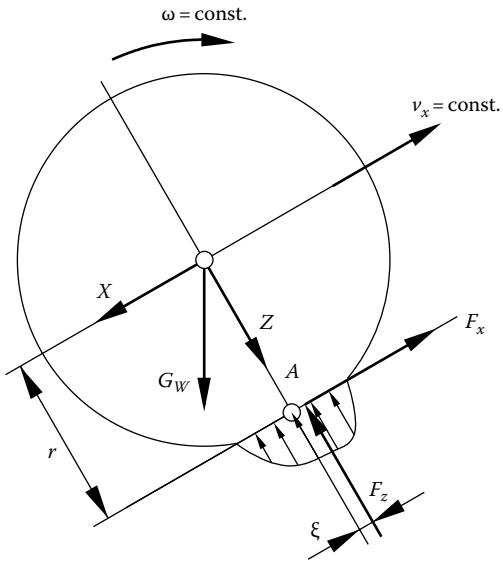


FIGURE 27.5 Tire moving with constant velocity without axle torques.

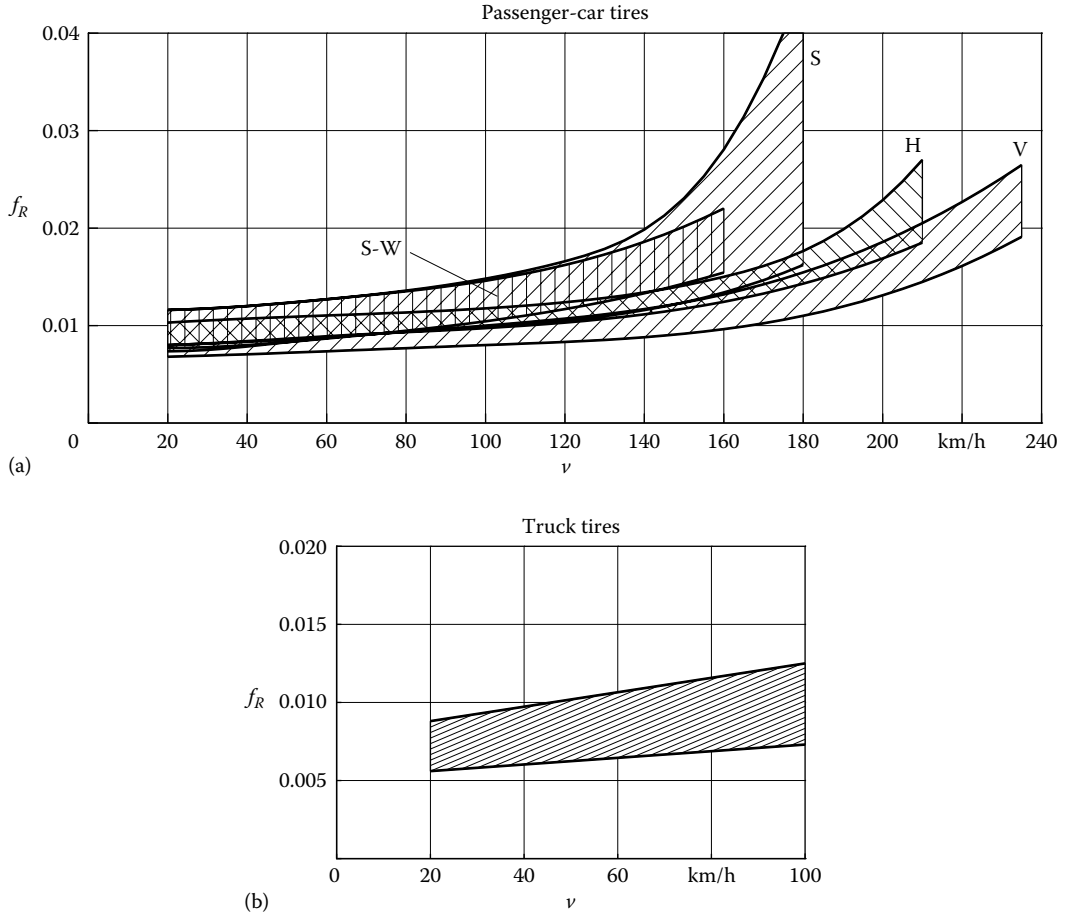


FIGURE 27.6 Rolling resistance coefficient for different types of (a) radial passenger-car tires (S-W: winter tires; permitted max. velocity for S: 180 km/h, H: 210 km/h, V: >210 km/h) and (b) truck tires (wheel loads between 20 and 30 kN). (From Mitschke, M. and Wallentowitz, H., *Dynamik der Kraftfahrzeuge*, Springer Verlag, Berlin, Germany, 2004.)

By stating the equation of equilibrium for the moments with respect to the wheel axle, the longitudinal tire force can be expressed by

$$F_x = -\frac{\xi}{r} F_z = -f_R F_z \tag{27.26}$$

The value of the rolling resistance coefficient f_R mainly depends not only on the type of tire and velocity (see Figure 27.6) but also on inflation pressure (see e.g., [7]) and tire temperature (see e.g., [6]). A water layer on the road surface also increases f_R (see e.g., [7]). Especially on soft soil, the rolling resistance will increase dramatically (see [14]).

27.2.3.2 Aerodynamic Phenomena and Drag

For many years, the aerodynamic flow around the moving vehicle has been intensively investigated; some essential findings and data can be found in Chapter 16 and [6,7,15,16].

The aerodynamic pressure distribution acting on the vehicle is represented by its reduction to a reference point and the corresponding forces and moments. Due to the complexity of this pressure system, mostly wind tunnel measurements are used to determine the aerodynamic coefficients used

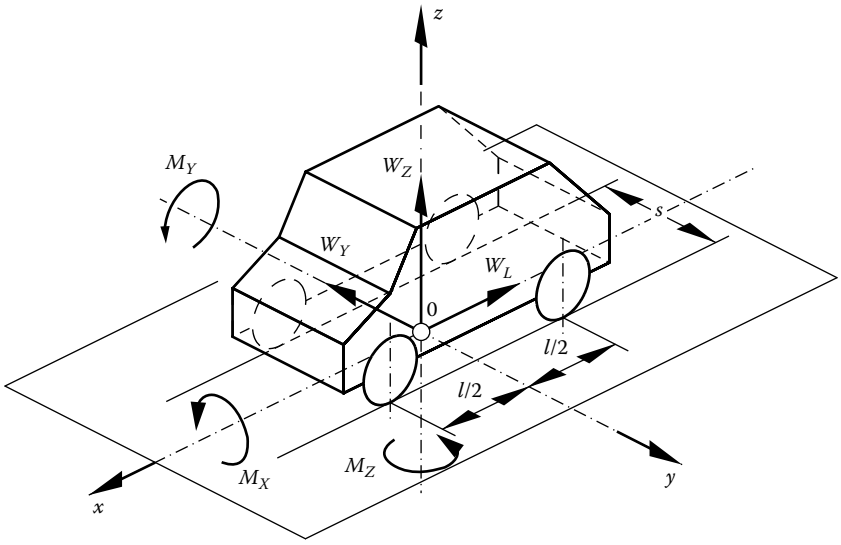


FIGURE 27.7 Aerodynamic forces and moments.

for the description of the reduction results for reference point 0 in the center of the rectangle made up by the contact points of the wheels (Figure 27.7 [7]).

Most relevant for the magnitude of forces and moments is the resultant air velocity vector $\vec{v}_R$, defined by its magnitude v_R and the aerodynamic side slip angle β_a (see Figure 27.8). The vector $\vec{v}_R$ results from the negative driving velocity of the vehicle $-\vec{v}$ and the velocity $\vec{v}_W$ of the ambient wind.

The dynamic pressure p_d for the vehicle is defined by (with air density ϱ)

$$p_d = \frac{\varrho v_R^2}{2} \tag{27.27}$$

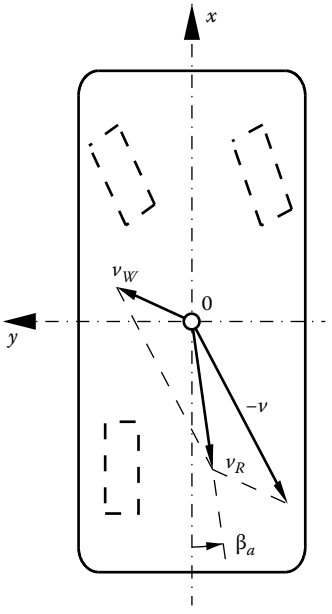


FIGURE 27.8 Resultant aerodynamic velocity $\vec{v}_R$ due to ambient wind $\vec{v}_W$ and vehicle velocity $\vec{v}$.

According to the *Vehicle Dynamics Terminology* (see [16,17]), the forces and moments acting on the vehicle at 0 (Figure 27.7) are described by applying force and moments coefficients $C_x \ldots C_{Mz}$

$$W_L = C_x A \cdot p_d, \quad W_Y = C_y A \cdot p_d, \quad W_Z = C_z A \cdot p_d$$
$$M_X = C_{Mx} A \cdot p_d l, \quad M_Y = C_{My} A \cdot p_d l, \quad M_Z = C_{Mz} A \cdot p_d l \tag{27.28}$$

with the vehicle area A (projected frontal area including tires and underbody parts). The coefficient C_x, responsible for the aerodynamic drag, is a specific value of the tangential aerodynamic coefficient $C_T(\beta_a)$, [15], $C_T(0) = C_x$. In general, for passenger cars, C_T is nearly constant for small aerodynamic side slip angles β_a, whereas a stronger influence can be seen for other vehicle body shapes (Figure 27.9).

C_x, $C_x A$ are often used to indirectly characterize the fuel consumption efficiency of cars and trucks. For today's passenger cars, values of $C_x = 0.3$ and $C_x A = 0.6$ m^2 are common, and for trucks, the values are about $C_x = 0.8$ and $C_x A = 5$–7 m^2 and even more, especially for trucks with trailers [4].

For small β_a, the coefficients C_Y, C_{Mx}, and C_{Mz} show a nearly linear increase with β_a. For example, there is a pronounced influence of the shape of the rear end (Figure 27.10); for more detailed characteristics, refer to [15]. Obviously, the box-type rear end of the station wagon has the advantage of a smaller aerodynamic yawing moments that can contribute to improved stability at higher velocities.

Nowadays, computational fluid dynamics (CFD) approaches are used to investigate aerodynamic details of the body shape as well [15], but generally the aerodynamic coefficients are still measured in wind tunnels.

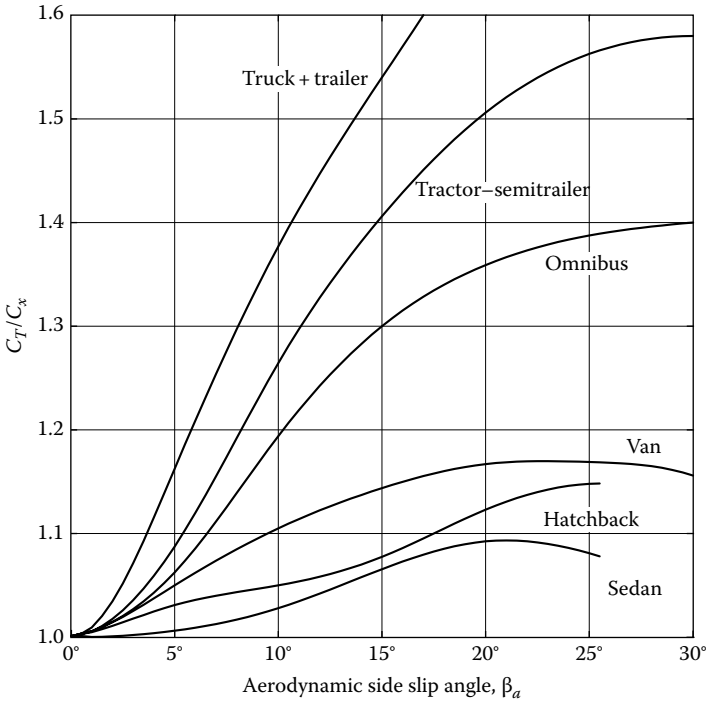


FIGURE 27.9 Influence of type of vehicle (vehicle body shape) on the normalized tangential aerodynamic coefficient $C_T(\beta_a)/C_x$.

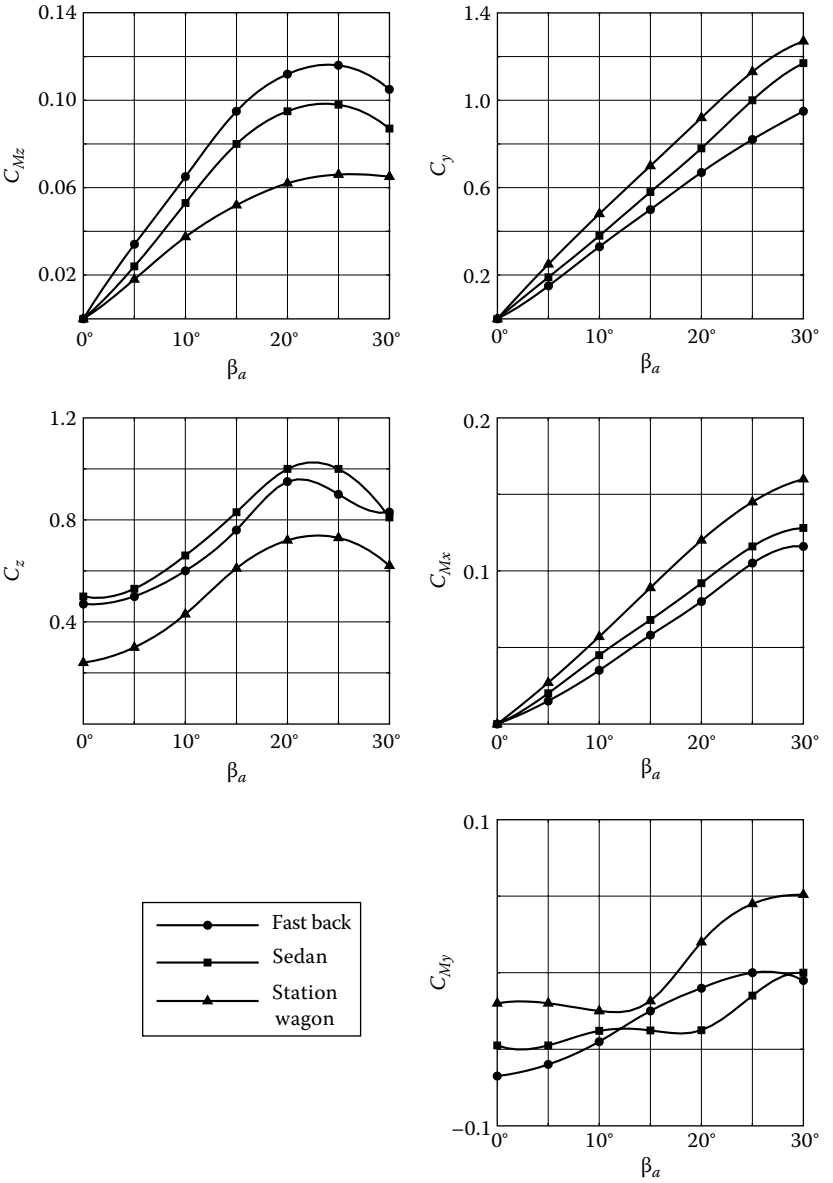


FIGURE 27.10 Aerodynamic coefficients as functions of aerodynamic side slip angle β_a for different rear-end shapes.

27.2.4 Vehicle Performance

Keeping in mind the description of the longitudinal dynamics in (27.17) and the driving resistances of W_{total} in (27.10), including rolling resistance and aerodynamic drag (Section 27.2.3), the engine torques $M_E(\omega_E)$, the overall gear ratio N, and the braking activities M_{Bi} will determine the performance of the vehicle.

27.2.4.1 Driving Performance

Here, no braking torques are applied and the engine torques is decisive for driving speed or acceleration performance (e.g., traction limit). The fuel consumption will not be considered in this context,

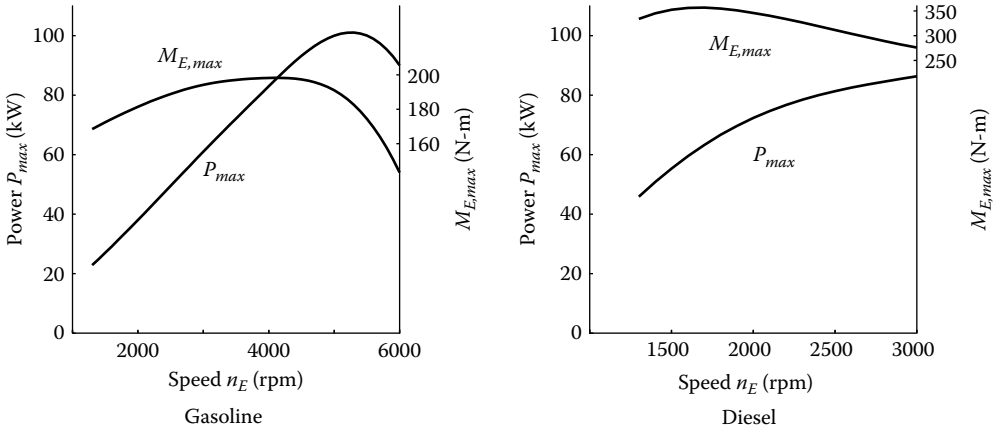


FIGURE 27.11 Examples for full-throttle engine characteristics, $M_{E,max}$, P_{max}.

but could be related to the engine characteristics [7,12]. Two typical full-throttle engine characteristics are shown in Figure 27.11 [6].

The power is defined by (engine speed n_E in r.p.m.)

$$P = M_E \cdot \omega_E = M_E \cdot \frac{n_E \pi}{30} \tag{27.29}$$

With throttle position $0 \le \lambda_T \le 1$, the applied torques M_E of the engine can be varied between engine drag $M_{E,d}$ and full-throttle engine torques $M_{E,max}$ in about a linear way:

$$M_E = (M_{E,max} - M_{E,d})\lambda_T + M_{E,d}$$

To obtain an overview of the vehicle performance for different speeds v_x and grades ϑ, the engine torques $M_E(\omega_E)$ is transformed to a net driving force $K_E = K_E(v_x, N_{Gn})$ as a function of v and different gear ratios N_{Gn}:

$$K_E = \frac{\eta M_E(\omega_E) \cdot N_{Gn} N_D}{r}, \quad \omega_E = v_x \cdot \frac{N_{Gn} N_D}{r}$$

With (27.17), (27.19), there follows

$$m\lambda a_x = K_E(v_x, N_{Gn}) - W_{total}(v_x, \vartheta) \tag{27.30}$$

Herein, the assumption is made again, that slip values remain small (and therefore also no friction limits are considered). Now, for example, with the resulting representative driving performance diagram in Figure 27.12, the maximum velocity v_{max} that can be accomplished on a horizontal road, $\vartheta = 0°$, with full throttle applied and in fourth gear, N_{G4}, can be pointed out from the intersection of the corresponding curve and the curve representing rolling and aerodynamic resistance, $W_{total} = f_R G + W_L$ (see (27.10)).

The figure also shows the net driving forces necessary for grades $q = -10\%$, 0%, $+10\%$ ($\vartheta = -5.7°$, $0°$, $+5.7°$). For example, with engine drag (no throttle applied), represented by $K_{E,d}$ and N_{G3}, a steady-state downhill velocity of v_1 is possible for $q = -10\%$. It should be mentioned that the engine drag $K_{E,d}$, especially at low velocities, strongly depends on the idle engine performance. The acceleration capability of the vehicle can be observed by the difference in K_E between the respective driving resistance curve (including the influence of grade) and the curve of the full net driving force $K_{E,max}(v_x, N_{Gn})$; for example, at velocity v_2, W_{total} with $q = 10\%$ and second gear N_{G2}, $m\lambda a_{x,2}$ is available for maximum acceleration. Driving conditions in between can be achieved by applying the corresponding throttle position.

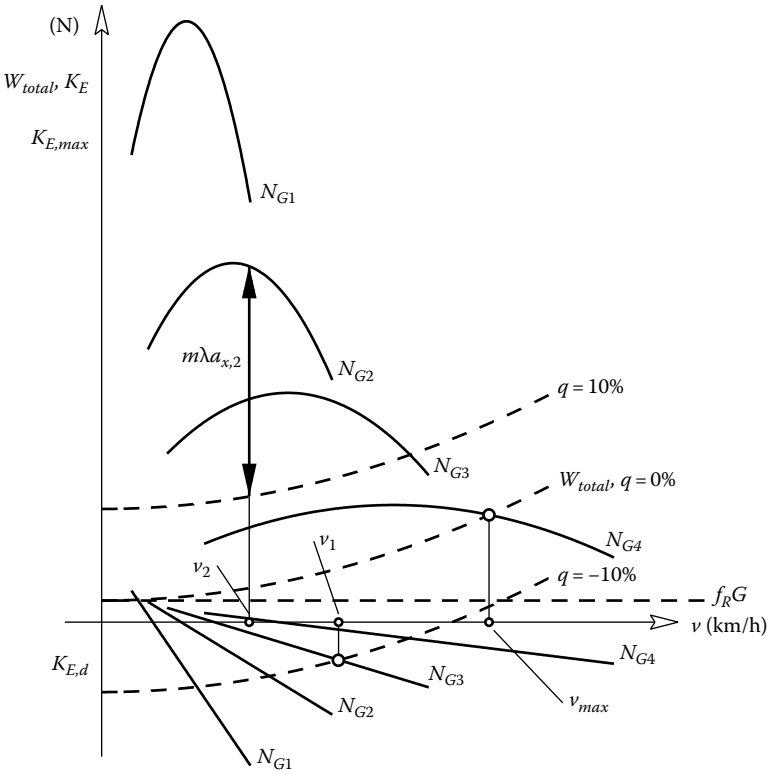


FIGURE 27.12 Driving performance diagram: principal characteristics for a passenger car with a gasoline engine; $K_{E,max}$, corresponding to maximum engine torques, are the upper four curves for the gears N_{Gn}; $K_{E,d}$ is the linear engine drag force at N_{Gn}.

The driving performance for starting at μ-split conditions (maximum longitudinal force coefficient, left side higher than/equal to right side: $\mu_{L,max} \geq \mu_{R,max}$) for different drive train configurations can be seen in Figure 27.13. Thereby, it is assumed that all wheels can work at their friction limits according to (27.20) and (27.23) through (27.25), still implying equal normal tire forces left and right. For a locked differential, the drive torques distribution is determined by the friction demands of the two wheels, while for the unlocked differential the friction limit of the wheel with the lower μ_{max} determines the longitudinal tire forces for both sides. For generalization, the dimensionless quantities a^*_{max}, W^* are introduced for the maximum acceleration (including grade and aerodynamic drag, see (27.25)) and the yaw moments due to the differences in longitudinal tire forces, see Figure 27.13,

$$W^* = \frac{(F_{x1} - F_{x2}) - (F_{x3} - F_{x4})}{Gl} \frac{s}{2} \tag{27.31}$$

The advantage of a higher traction capability for locked axle differentials at μ-split leads to increasing yaw moments due to different longitudinal forces at the left and right track. The induced yaw moments W^* has to be compensated by the driver's steering effort. For four-wheel drive (A), front-wheel drive (FD), and rear-wheel drive (RD), all without locked differentials, always the low friction side, respectively the corresponding longitudinal force, is responsible not only for the maximum possible a^*_{max} but also for $W^* = 0$, and principally no correction by the driver is necessary.

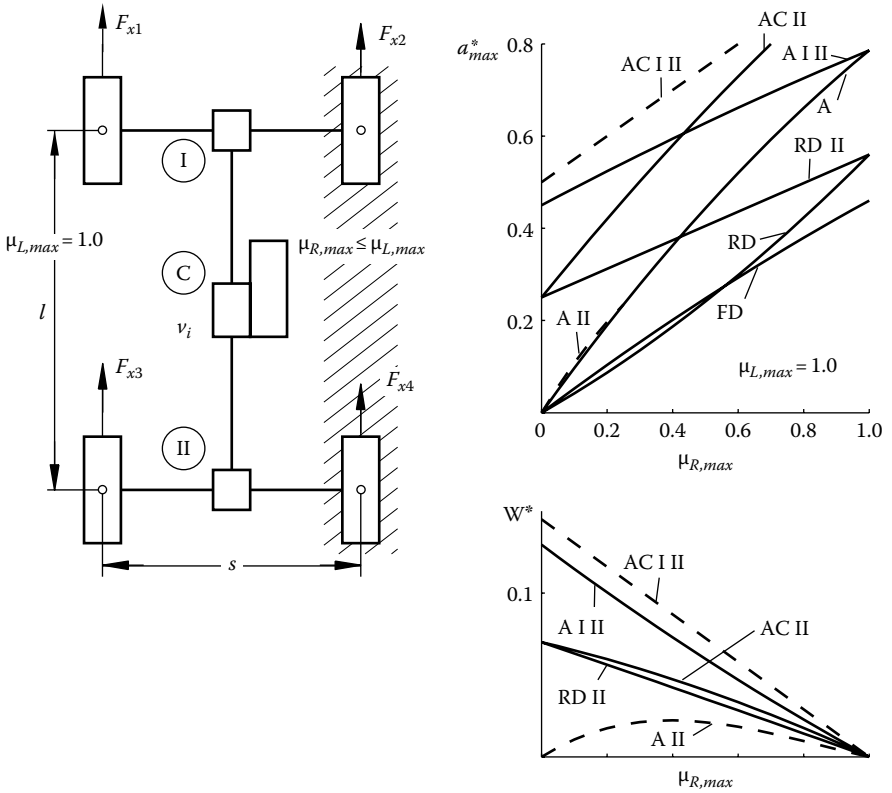


FIGURE 27.13 Starting on μ-split: Maximum possible g-normalized acceleration a^*_{max} and (necessary) normalized yaw moments W^* for rear-wheel drive (RD), front-wheel drive (FD), and four-wheel drive (A) with $\nu_F = \nu_R = 0.5$; C = locked central differential; I/II = locked axle differential front/rear.

Moreover, some more effects of a four-wheel drive vehicle can be noticed: if only one axle differential is locked (e.g., A II), there will be no real improvement in traction capability due to the interaction with the wheels at the unlocked axle; again the low friction side determines a^*_{max}, but there is still a noticeable W^*! Of course, the highest utilization of the available friction can be achieved with all differentials locked, AC I II; however, this configuration induces the largest interfering yaw moments.

27.2.4.2 Braking

In the following, "braking" is considered as a period of constant deceleration. The essential equations to represent braking dynamics have already been established in (27.2) and (27.23) through (27.25). Introducing dimensionless quantities and the effective normalized deceleration β (for small grade angles ϑ)

$$\frac{2F_{xi}}{G} = \bar{F}_{xi}, \quad \frac{h}{l} = \bar{h}$$
$$\frac{2F_{zi}}{G} = \bar{F}_{zi}, \quad \frac{l_i}{l} = \bar{l}_i, \quad i = F, R \tag{27.32}$$

$$a^* = \left(\frac{a_x}{g} + \vartheta + \frac{W_L}{G} \right), \quad \beta = -a^* > 0 \tag{27.33}$$

Equations (27.2) and (27.23) through (27.24) read

$$-\beta = \bar{F}_{xF} + \bar{F}_{xR} \tag{27.34}$$

$$\bar{F}_{zF} = \bar{l}_R + \beta\bar{h} \tag{27.35}$$

$$\bar{F}_{zR} = \bar{l}_F - \beta\bar{h} \tag{27.36}$$

and consequently

$$\bar{F}_{zF} + \bar{F}_{zR} = 1 \tag{27.37}$$

In these equations, small terms are neglected once again, see Section 27.2.2. It is assumed further that the vehicle and the environmental conditions (μ-conditions) are symmetric for the vehicle's longitudinal central plane.

For maximum deceleration, both the front and rear wheels need to utilize the existing friction conditions to their full extent at the same time. With equal conditions at both axles,

$$\mu_{F,max} = \mu_{R,max} = \mu_{max} \tag{27.38}$$

the braking forces are (braking means $\bar{F}_{xi} < 0$ ($i = F, R$))

$$\bar{F}_{xR} = -\mu_{max}\bar{F}_{zR}, \quad \bar{F}_{xF} = -\mu_{max}\bar{F}_{zF} \tag{27.39}$$

and by considering (27.37), (27.34) yields the maximum deceleration

$$\beta = \beta_{ideal} = \mu_{max} \tag{27.40}$$

and, with (27.35) and (27.36), the necessary braking force balance

$$\frac{-\bar{F}_{xR}}{-\bar{F}_{xF}} = \frac{\bar{l}_F - \bar{h}\mu_{max}}{\bar{l}_R + \bar{h}\mu_{max}} \tag{27.41}$$

It becomes evident that for each braking traction coefficient μ_{max}, another (ideal) braking force balance would be required for maximum deceleration (see Figure 27.14, [18,19]).

The diagonal lines in Figure 27.14, according to (27.34), represent all possible combinations of $-\bar{F}_{xR}$ and $-\bar{F}_{xF}$ that result in the same normalized deceleration β. The curved line is the selection of the points of the ideal braking force balance in (27.41), with regard to (27.38), with both braking forces utilizing μ_{max}.

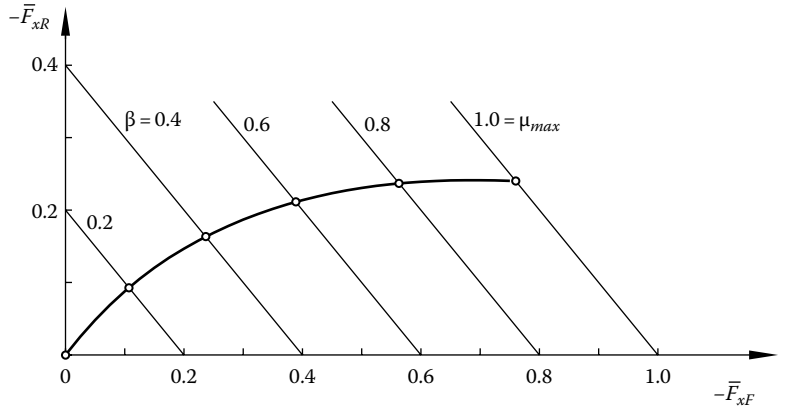


FIGURE 27.14 Ideal braking force balance for $\bar{l}_F = 0.52$, $\bar{h} = 0.28$.

If only the front wheels are operating at limit conditions

$$(-\bar{F}_{xF}) = \mu_{max}\bar{F}_{zF} \tag{27.42}$$

the rear braking force can be derived by using (27.34) and (27.35) to

$$(-\bar{F}_{xR}) = (-\bar{F}_{xF})\frac{1-\mu_{max}\bar{h}}{\mu_{max}\bar{h}} - \frac{\bar{l}_R}{\bar{h}} \tag{27.43}$$

Analogously, in case only the rear wheels are at the limit

$$(-\bar{F}_{xR}) = \mu_{max}F_{zR} \tag{27.44}$$

the front braking force is

$$(-\bar{F}_{xF}) = -(-\bar{F}_{xR})\frac{1+\mu_{max}\bar{h}}{\mu_{max}\bar{h}} + \frac{\bar{l}_F}{\bar{h}} \tag{27.45}$$

These two relations, (27.43) and (27.45), limit the applicable area for braking (see straight "limit" lines in Figure 27.15). Larger normalized braking forces are not possible due to the locking of either the front or the rear wheels.

If a constant braking torques balance between front and rear wheels is implemented in the braking system, with the braking forces approximately proportional to the braking torques (if slip changes are not too large), the following relation holds:

$$\frac{-\bar{F}_{xF}}{-\bar{F}_{xR}} = k \tag{27.46}$$

which corresponds to the straight line marked with (*c*) in Figure 27.15. With a specific set of the braking balance, for example, represented by point *A* on this line, the related value of the normalized deceleration β can be found by the intersection of the −45°-inclined line with the horizontal axis.

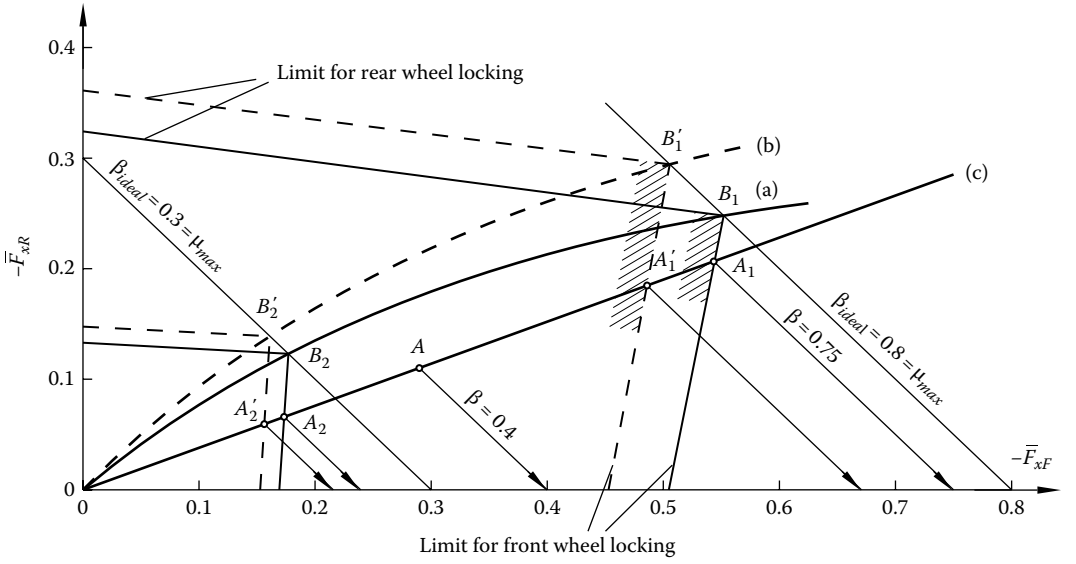


FIGURE 27.15 Braking force distribution diagram (with normalized braking forces $-\bar{F}_{xF}$, $-\bar{F}_{xR}$) for a passenger car for two loading conditions (curve (a): $\bar{l}_F = 0.47$, $\bar{h} = 0.2$; curve (b): $\bar{l}_F = 0.52$, $\bar{h} = 0.19$) and two braking traction coefficients $\mu_{max} = 0.3$; 0.8 with constant braking force balance (curve (c)): $-\bar{F}_{xF}/-\bar{F}_{xR} = k = 2.63$.

With the braking force distribution diagram of Figure 27.15, the performance of the braking system, including its limits due to friction conditions, can be demonstrated. The curves (a) and (b) represent the ideal braking force balance for two different loading conditions and therefore points B_1', B_1 and B_2', B_2 correspond to ideal tuning for the (normalized) braking forces. The implemented braking system (represented by line (c)) is set in such a way that with increasing braking pressure—that means going along line (c)—the front wheels will lock first, marked with A_1', A_1 and A_2', A_2. Looking on the possible normalized decelerations β, it becomes obvious that only less than the ideal deceleration $\beta_{ideal} = \mu_{max}$ can be achieved. Braking systems are typically designed in such a way that front wheels will lock first in order to avoid the spin-out of the vehicle at a critical braking maneuver—a vehicle with locked front wheels will usually move straight ahead. In this case, the loss of steering capability of the front wheels—due to a reduction in lateral forces for locked wheels—can be improved with an ABS! The shaded areas in Figure 27.15 indicate a possible working area for the ABS for $\mu_{max} = 0.8$. It does not enhance the possible deceleration substantially but avoids the locking of the wheels. Within the shaded areas, the ABSs aim to reach ideal braking. A detailed description of an ABS (and ESP) is presented in Chapter 30.

Besides the design of the braking system with constant $-\bar{F}_{xF}/-\bar{F}_{xR} = k$, there are various other possibilities, for example, a piecewise linear relation that can be closer to the ideal curves (a), (b) and therefore utilizes the available friction limits better (e.g., [12]).

27.2.5 Reaction, Braking Procedure, and Stopping

After the driver has recognized and decided that a braking action has to be executed, some time elapses until the deceleration of the vehicle becomes effective. This period is called prebraking time, t_{pre}, which is typically about 0.8–1.0 s [7,12], and consists of the reaction, transfer, and response time. For simplicity, half of the brake pressure build-up time, t_s, is usually included here as well.

The reaction time is the time between the defined incitement to action and the start of the first specifically targeted action, 0.2–0.3 s [12]. Then the driver requires about 0.2 s [8] to transfer the foot from the gas to the brake pedal (transfer time). Some more time elapses, called the response time, approximately 0.05 s [7], to transmit pressure from the brake pedal to the braking system and the braking action becomes effective. This period, $(t_1 - t_0)$, is illustrated in Figure 27.16.

The brake pressure is built up until t_1' and full deceleration becomes effective at t_2, $t_s = t_2 - t_1$, approximately 0.17 s [7].

After the "mean maximum retardation" period, $(t_3 - t_2)$, the brake release ends the braking operation at t_4. Thus, the interval, $(t_4 - t_1)$, is the total braking time t_B. As mentioned earlier, half of the

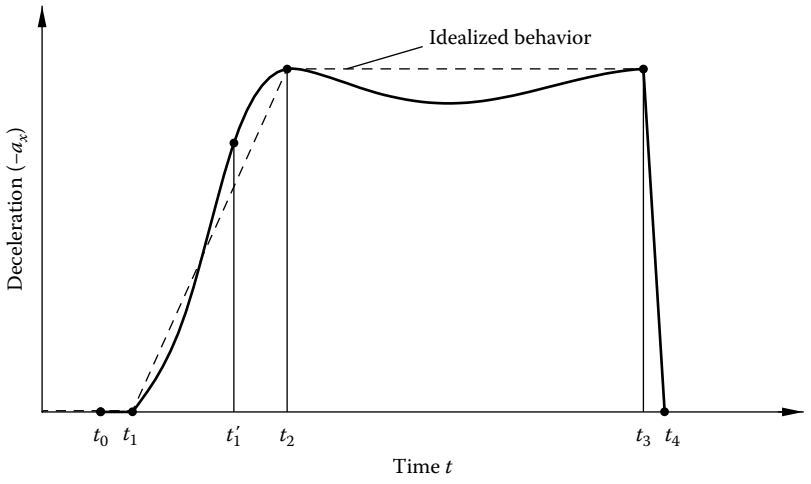


FIGURE 27.16 Characteristics of a braking procedure.

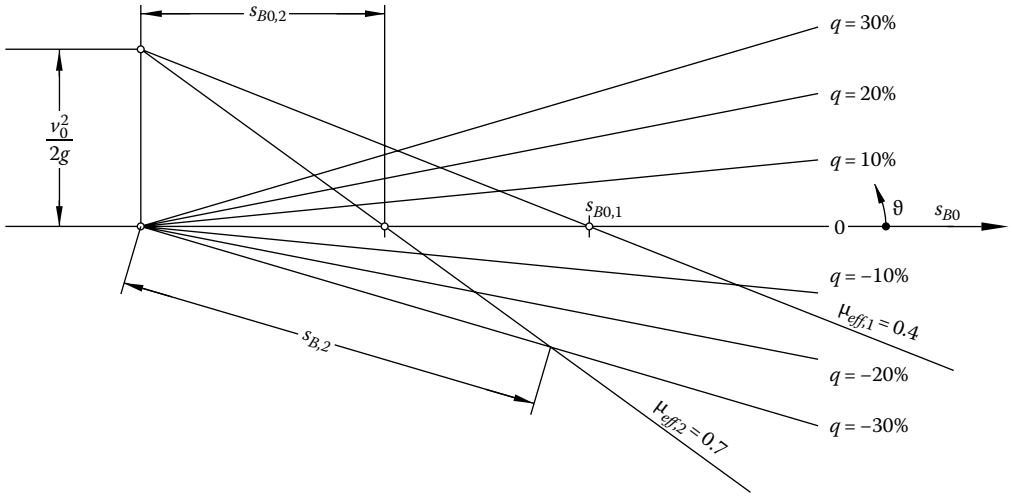


FIGURE 27.17 Braking distance s_{Bi} to standstill as a function of grade ϑ (for constant deceleration).

brake pressure build-up time, $t_s/2$, is usually drawn off to count for the prebraking time. Moreover, often the decline time, $(t_4 - t_3)$, is neglected compared to $(t_3 - t_2)$. An idealized braking behavior, which maps the time-dependent deceleration in a linear way, is plotted in Figure 27.16.

Finally, the stopping time of the vehicle is the sum of the prebraking time, t_{pre}, and the braking time, t_B. Consequently, the vehicle passes the distance $t_{pre} \cdot v$ during the first period, and together with the braking distance s_B during t_B, the full stopping distance.

For a time-varying deceleration, a time integration of the equations of motion is required in general, and, may be additionally, the locking of wheels needs to be taken into account. Nevertheless, as stated at the beginning of Section 27.2.4.2, many fundamental considerations are based on the assumption of a constant deceleration. In this way, ($\beta = -a_x = \text{const.} > 0$), the braking distance, s_B, can be estimated and derived by using (27.2) and (27.3) for different grades ϑ and no locking wheels, with an effective braking force coefficient

$$\mu_{eff} = \left| \frac{-(2F_{xF} + 2F_{xR})}{2F_{zF} + 2F_{zR}} \right| = \frac{m\beta - W_L - mg\sin\vartheta}{mg\cos\vartheta - W_Z} \tag{27.47}$$

Disregarding the aerodynamic effects ($W_L \approx 0$, $W_Z \approx 0$) with respect to the other quantities, the braking distance from the initial velocity v_0 to standstill is

$$s_B = \frac{v_0^2}{2g(\mu_{eff}\cos\vartheta + \sin\vartheta)}, \quad \text{for } \vartheta = 0: \quad s_{B0} = \frac{v_0^2}{2g\mu_{eff}} \tag{27.48}$$

Figure 27.17 illustrates for two different values of $\mu_{eff,j}$, the effect of different grades on the braking distance to standstill with respect to a horizontal road; thus, for $\mu_{eff,2} = 0.7$, the braking distance $s_{B,2}$ for a grade of $q = -30\%$ is nearly twice as long as the braking distance $s_{B0,2}$ on a horizontal road.

27.2.6 State-Space Description, Inverse Dynamics, and Basics for Longitudinal Control

By introducing simple linear characteristics for the determination of the applied longitudinal tire forces

$$F_{xi} = C_{si}s_i \quad i = F, R \tag{27.49}$$

with the slip stiffness C_{si} and the longitudinal slip (rolling radius r_i)

$$s_i = \frac{r_i\omega_i - v_x}{v_x} \quad i = F, R \tag{27.50}$$

and by recalling the equations of motion for the vehicle (27.2) through (27.4) and the simplified Euler's equations for the front and rear wheels (see (27.5), $M_{Fi} = 0$, $M_{Ai} = M_{Di} - M_{Bi}$ (i=F, R)) with the applied torques M_{Ai},

$$I_F\dot{\omega}_F = M_{AF} - f_R F_{zR} r_F - F_{xF} r_F \tag{27.51}$$

$$I_R\dot{\omega}_R = M_{AR} - f_R F_{zR} r_R - F_{xR} r_R \tag{27.52}$$

the equations of motion for the system can be stated in the form

$$\begin{gathered} \dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}) \\ \underline{x} = [v_x, \omega_F, \omega_R]^{\mathrm{T}}, \quad \underline{u} = [M_{AF}, M_{AR}]^{\mathrm{T}} \end{gathered} \tag{27.53}$$

with state-vector $\underline{x}$ and input vector $\underline{u}$. Setting $M_{AR} = 0$ or $M_{AF} = 0$ defines front- or rear-wheel drive, with a fixed relation M_{AR}/M_{AF}, both four-wheel drive and braking with a constant braking force balance can be taken into account.

By solving (27.53), which is generally a set of nonlinear differential equations, numerically, the vehicle performance can be investigated also for nonconstant deceleration, which has been assumed, for example, in Section 27.2.4.2. Also, locked wheels can be considered by $-F_{xi} = \mu_s F_{zi}$.

For given $v(t)$, for example, as a result from a driving test, the required torques M_{AF}, M_{AR} can be derived by solving (27.53) with respect to $\underline{u}$ ("inverse dynamics").

To linearize (27.53) with respect to a given $v_g = v_g(t)$, see [20]; here, quasi steady-state values v_s are used (and thus only moderate accelerations are presumed), and for horizontal roads,

$$v_x = v_s + \Delta v_x \tag{27.54}$$

$$\omega_i = \omega_{i,s} + \Delta\omega_i, \quad i = F, R \tag{27.55}$$

$$M_{Ai} = M_{Ai,s} + \Delta M_{Ai}, \quad i = F, R \tag{27.56}$$

Similar assumptions are made to derive (27.23) and (27.24); moreover, the normal tire forces are set constant for the determination of the rolling resistance, and the influence on C_{si} in (27.49) is neglected. The aerodynamic drag in (27.27) and (27.28) is included (for small longitudinal ambient wind $v_W \ll v_s$, $k_x = C_x A \varrho/2$) by

$$W_L = k_x(v_s + \Delta v_x + v_W)^2 \approx k_x v_s^2 + 2k_x v_s \Delta v + 2k_x v_s v_W \tag{27.57}$$

The linearized relation for (27.50) reads

$$s_i \approx \frac{r_i\omega_{i,s} - v_s}{v_s} + \frac{r_i\Delta\omega_i}{v_s} - \frac{\Delta v_x}{v_s} \quad i = F, R \tag{27.58}$$

when small longitudinal slip values with respect to steady-state conditions are assumed, which means $(r_i\omega_{i,s} - v_s)/v_s \ll 1$ or $r_i\omega_{i,s} \approx v_s$. For simplification, $r_F = r_R = r$.

The steady-state values for $M_{Ai,s}$ can be derived from $\underline{0} = \underline{f}(\underline{x}_s, \underline{u}_s)$ and balance rolling resistance, aerodynamic drag, and grade (here $\vartheta_{F,s} = 0$). Finally, the linearized system equations for the deviations from steady state are in state-space notation:

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} + \underline{g}_w w \tag{27.59}$$

with

$$\underline{x} = \begin{bmatrix} \Delta v_x \\ \Delta\omega_F \\ \Delta\omega_R \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -\dfrac{2\left(C_{sF} + C_{sR} + k_x v_s^2\right)}{m v_s} & \dfrac{2rC_{sF}}{m v_s} & \dfrac{2rC_{sR}}{m v_s} \\ \dfrac{rC_{sF}}{I_F v_s} & -\dfrac{r^2 C_{sF}}{I_F v_s} & 0 \\ \dfrac{rC_{sR}}{I_R v_s} & 0 & -\dfrac{r^2 C_{sR}}{I_R v_s} \end{bmatrix}$$

$$\underline{u} = \begin{bmatrix} \Delta M_{AF} \\ \Delta M_{AR} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ \dfrac{1}{I_F} & 0 \\ 0 & \dfrac{1}{I_R} \end{bmatrix}, \quad \underline{g}_w = \begin{bmatrix} \dfrac{1}{m} \\ 0 \\ 0 \end{bmatrix}, \quad w = -2k_x v_s v_w - mg\Delta\vartheta_F$$

Stability of (27.59) can be analyzed by calculating the eigenvalues λ_i of the system matrix $\mathbf{A}$ from solving the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$. In a symbolic way, the asymptotic stability of the linearized system (and as a consequence also locally of the nonlinear system) can be proved by applying Hurwitz' stability theorem. As the stability margin is quite small, (initial) disturbances vanish only slowly and controlled dynamics may be advisable. This can be achieved by applying an input vector $\underline{u} = -\mathbf{K}\underline{x}$, where $\mathbf{K}$ can easily be determined with the afore-listed matrices and elementary control theory methods (see Chapter 6). As constant disturbances still will not decay, a PI-state controller may be recommended.

27.3 LATERAL VEHICLE BEHAVIOR

27.3.1 Application and Vehicle Model Features

A simple model for basic analysis of lateral vehicle behavior (for more advanced models and analysis see Chapter 28) is the (linearized) two-wheel vehicle model in Figure 27.18 (e.g., [1,2,5,6,7,8,9,10,20]), where the two wheels of an axle are substituted by one wheel in the vehicle's central plane. The resulting two substitutive wheels for the two axles of a passenger car are assumed to be included into the vehicle mass and inertia (neglecting inertia effects due to wheel spin), and comprise not only the properties of the tires but also those of the wheel suspensions.

The limits for the validity of the model, for example, with respect to the lateral acceleration, and what kind of applications can be studied, depend on the additional model characteristics. In general, investigations based on a two-wheel (or a simplified four-wheel) vehicle model are as follows:

1. Steady-state cornering, steering behavior, and characterization by standardized quantities (e.g., understeer gradient)
2. Frequency response
3. Stability analysis

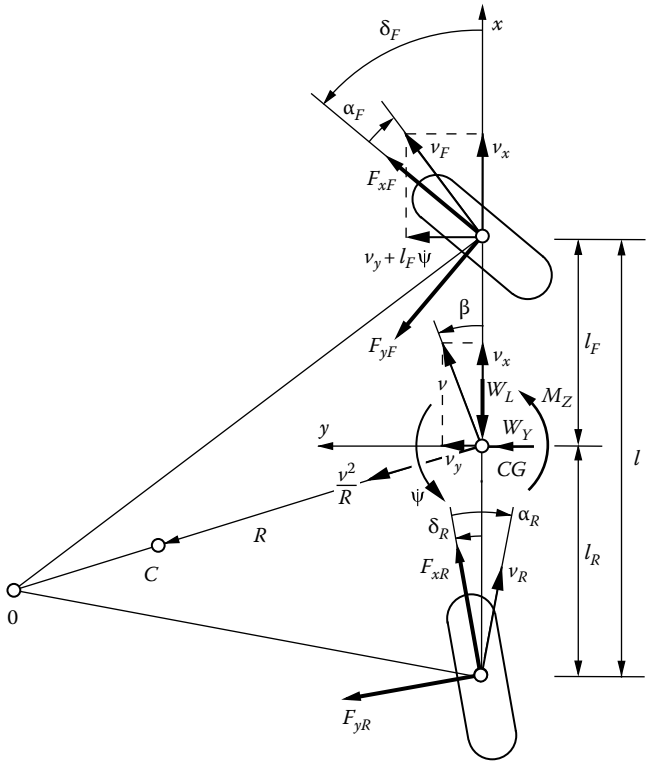


FIGURE 27.18 Two-wheel vehicle model (with possible rear-wheel steering).

4. Open-loop simulations with given steering angle and velocity (profile), closed-loop simulations, for example, driving along a given trajectory
5. Control model for design of driver-assistance systems or driver models
6. Basic influences of vehicle and suspension design
7. Identification of essential vehicle parameters based on measurements

Typically, for applications 1–5, the vehicle moves on a horizontal flat surface. Further, the velocity v_x is often assumed constant and the front-wheel steering angle δ_F (with rear-wheel steering $\delta_R = 0$ or $\delta_R = \delta_R(\delta_F, v_x, \ldots)$) is considered an input variable.

For applications 6 and 7, the model is extended by including the vehicle body motion, especially the roll motion based on the assumption of a fixed roll axis, and/or sometimes nonlinear tire characteristics as well.

27.3.2 Linear Two-Wheel Vehicle Model

In English literature, the two-wheel vehicle model often referred to as "bicycle model" or "single track model," but both terms can be misunderstood in the narrow sense of the word. In the following, a basic understanding and analysis of the (in most cases) linear two-wheel vehicle model is presented.

27.3.2.1 Equations of Motion

To explain the principal lateral behavior of a two-axle vehicle, the two-wheel vehicle model is used in any basic lectures of vehicle dynamics. It is based on the model shown in Figure 27.18, linearized with respect to an undisturbed straight ahead driving with constant longitudinal velocity v_x.

Thus, angles δ_F, δ_R, α_F, α_R, and β are assumed to be small. Then, the linearized equations of motion are

$$ma_x = (F_{xF} - F_{yF}\delta_F) + (F_{xR} - F_{yF}\delta_R) - W_L \tag{27.60}$$

$$ma_y = (F_{xF}\delta_F + F_{yF}) + (F_{xR}\delta_R + F_{yR}) + W_Y \tag{27.61}$$

$$I_Z\ddot{\psi} = (F_{xF}\delta_F + F_{yF})l_F - (F_{xR}\delta_R + F_{yR})l_R + M_Z \tag{27.62}$$

with the total vehicle mass m and moments of inertia I_Z with respect to the z-axis at the center of gravity CG. The yaw rate $r = \dot{\psi} = \dot{\psi}(t)$ describes the rotational motion of the vehicle (fixed frame x–y–z) with respect to the inertial frame. Note that the instantaneous center 0 of rotation/zero velocity of the vehicle and the center C of curvature ($1/R$) of the trajectory of CG do not coincide in the general case, but for steady-state cornering! Presuming that the longitudinal tire forces remain small, the terms $F_{xi}\delta_i$ ($i = F, R$) in (27.61) and (27.62) can also be neglected and (27.60) is no longer relevant for the lateral vehicle motion, reducing the equations of motion for neglected aerodynamics, $M_Z = 0$, $W_L = 0$, $W_Y = 0$, to

$$ma_y = F_{yF} + F_{yR} \tag{27.63}$$

$$I_z\ddot{\psi} = F_{yF}l_F - F_{yR}l_R \tag{27.64}$$

For a given velocity, the longitudinal tire forces F_{xi} are constraint forces. They can be obtained from (27.60), with a_x, for example, from (27.67), considering front-, rear-, or all-wheel drive, respective, and the rolling resistance. The lateral tire force F_{yi} of a substitutive tire in (27.63) and (27.64) is described by a linear approach using the side slip angle α_i of the respective axle and the effective cornering stiffness C_i:

$$F_{yi} = C_i\alpha_i, \quad i = R, F \tag{27.65}$$

Looking at the kinematics of the vehicle and considering the yaw angular velocity $\dot{\psi}$ of the body fixed frame x–y–z to the inertial system, yields for the accelerations

$$a_y = \dot{v}_y + \dot{\psi}v_x \tag{27.66}$$

$$a_x = -\dot{\psi}v_y \tag{27.67}$$

and for the linearized side slip angles

$$\alpha_F = \delta_F - \frac{v_y + l_F\dot{\psi}}{v_x}, \quad \alpha_R = \delta_R - \frac{v_y - l_R\dot{\psi}}{v_x} \tag{27.68}$$

Alternatively to the 2 degrees of freedom (DOFs) v_y and $\dot{\psi}$, the lateral velocity v_y of the vehicle can be replaced by the side slip angle β of the vehicle,

$$\beta = \frac{v_y}{v_x}, \quad \dot{\beta} = \frac{\dot{v}_y}{v_x} \tag{27.69}$$

leading to

$$a_y = v_x(\dot{\beta} + \dot{\psi}) \tag{27.70}$$

$$\alpha_F = \delta_F - \beta - \frac{l_F\dot{\psi}}{v_x}, \quad \alpha_R = \delta_R - \beta + \frac{l_R\dot{\psi}}{v_x} \tag{27.71}$$

After incorporating the kinematic relations (27.66) and (27.68) and the lateral tire force description (27.65) in the equations of motion (27.63) and (27.64), the set of linear differential equations read in state-space notation

$$\dot{\underline{x}} = \mathbf{F}\underline{x} + \mathbf{G}\underline{\delta} \tag{27.72}$$

with the state vector $\underline{x} = [v_y,\ \dot{\psi}]^T = [v_y,\ r]^T$, the input vector $\underline{\delta} = [\delta_F,\ \delta_R]^T$, the system matrix

$$\mathbf{F} = \begin{bmatrix} -\dfrac{C_F + C_R}{mv_x} & -\dfrac{(l_F C_F - l_R C_R)}{mv_x} - v_x \\ -\dfrac{\left(l_F C_F - l_R C_R\right)}{I_Z v_x} & -\dfrac{l_F^2 C_F + l_R^2 C_R}{I_Z v_x} \end{bmatrix} \tag{27.73}$$

and the input matrix

$$\mathbf{G} = \begin{bmatrix} \dfrac{C_F}{m} & \dfrac{C_R}{m} \\ \dfrac{l_F C_F}{I_Z} & \dfrac{l_R C_R}{I_Z} \end{bmatrix} \tag{27.74}$$

The rear-wheel steering angle δ_R may be included for investigations with respect to a four-wheel steering or other steering effects due to suspension characteristics.

Equation in (27.72), with $\delta_R = 0$, can also be written as two linear second-order differential equations for the yaw rate r and side slip angle β of the vehicle by applying (27.69) and a separation of variables [21],

$$\ddot{\beta} + 2K_1\dot{\beta} + K_2\beta = \frac{C_F}{mv_x}\dot{\delta}_F - \frac{C_F\left(l_F m v_x^2 - C_R l_R l\right)}{I_Z m v_x^2}\delta_F \tag{27.75}$$

$$\ddot{r} + 2K_1\dot{r} + K_2 r = \frac{l_F C_F}{I_Z}\dot{\delta}_F + \frac{C_F C_R l}{I_Z m v_x}\delta_F \tag{27.76}$$

with

$$K_1 = \frac{I_Z(C_F + C_R) + m\left(C_F l_F^2 + C_R l_R^2\right)}{2I_Z m v_x} \tag{27.77}$$

$$K_2 = \frac{l^2 C_F C_R + (C_R l_R - C_F l_F) m v_x^2}{I_Z m v_x^2} \tag{27.78}$$

This form may then be used to easily derive transfer functions to switch from time to frequency domain or, for instance, for the analysis of vehicle stability with respect to parameter influences.

27.3.2.2 Steady-State Cornering

For comparability reasons with the standard definitions and interpretations, [6,17], a front-steered vehicle is considered further on only ($\delta_R = 0$).

At steady-state cornering with radius R and trim conditions,

$$\beta_{st} = \text{constant}, \quad r_{st} = \text{constant}, \quad a_{y,st} = \frac{v_x^2}{R} \tag{27.79}$$

the equations of motion (27.75) through (27.78) result in

$$\delta_{F,st} = \delta_{F0} + \frac{C_R l_R - C_F l_F}{C_R C_F l} m a_{y,st} \tag{27.80}$$

$$\beta_{st} = \beta_0 - \frac{l_F}{C_R l} m a_{y,st} \tag{27.81}$$

with the front-wheel steering angle δ_{Fo} (Ackermann steering angle δ_a) and side slip angle β_0 of the vehicle for trim $v_x \approx v \to 0$, according to Figure 27.19. Since ideally in this case no lateral tire forces are acting at the substitutive tires and therefore also no tire side slip angles, the angles β_0 and $\delta_{F0} = \delta_a$ may be calculated directly from the geometric situation of the vehicle. For small angles, one yields

$$\beta_0 = \frac{l_R}{R} \tag{27.82}$$

$$\delta_a = \delta_{F0} = \frac{l}{R} \tag{27.83}$$

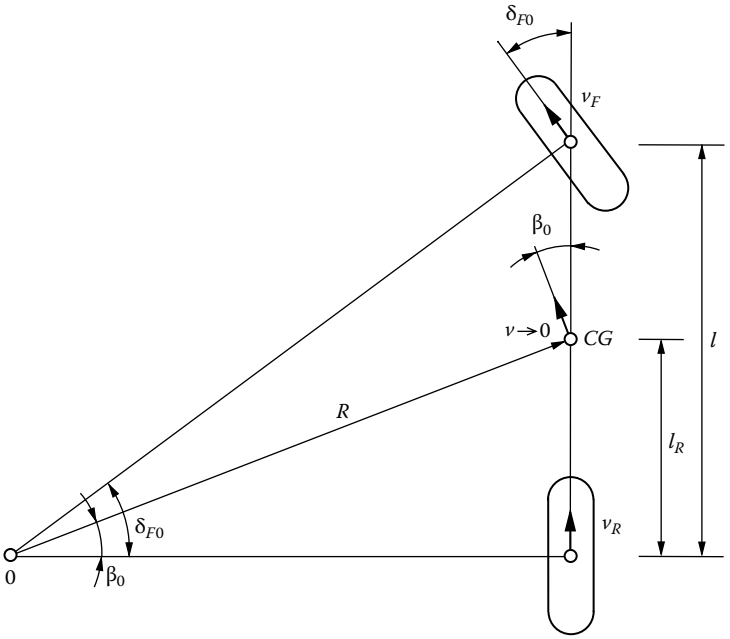


FIGURE 27.19 Steady-state cornering for $v \to 0$.

or after eliminating R

$$\delta_a = \frac{l}{v_x^2} a_{y,st} \tag{27.84}$$

In some standard definitions, the steering angle of the hand wheel δ_H appears as well and follows from the overall steering ratio i_s

$$\delta_H = i_s \delta_F \tag{27.85}$$

A very useful quantity to describe handling and steering behavior is the understeer/oversteer gradient K_{US} (see also SAE J670e, [6,17], or DIN 70 000 for "Eigenlenkgradient" $EG = K_{US}/l$), which is to be obtained by subtracting the Ackermann steering angle gradient from the ratio of the steering wheel angle gradient to the overall steering ratio, that is,

$$K_{US} = \frac{1}{i_s} \frac{\mathrm{d}\delta_{H,st}}{\mathrm{d}a_{y,st}} - \frac{\mathrm{d}\delta_a}{\mathrm{d}a_{y,st}} \tag{27.86}$$

Combining (27.86) and (27.80) gives the steering angle for the linear system

$$\frac{\delta_{H,st}}{i_s} = \delta_{F,st} = \delta_a + K_{US} a_{y,st} \tag{27.87}$$

with the understeering gradient

$$K_{US} = \frac{m(C_R l_R - C_F l_F)}{l C_R C_F} \tag{27.88}$$

The steering/handling properties are then characterized by the value of K_{US}:

$$\begin{aligned} K_{US} &> 0 \quad \text{understeer} \\ K_{US} &= 0 \quad \text{neutral steer} \\ K_{US} &< 0 \quad \text{oversteer} \end{aligned} \tag{27.89}$$

Another way to interpret the steering behavior makes use of the difference of the side slip angles of the respective substitutive wheels. According to Figure 27.20, there follows for small angles

$$\delta_{F,st} = \frac{l}{R} + (\alpha_F - \alpha_R) = \delta_a + (\alpha_F - \alpha_R) \tag{27.90}$$

$$\beta_{st} = \frac{l_R}{R} - \alpha_R = \beta_0 - \alpha_R \tag{27.91}$$

From (27.90), it can be seen that the radius R of the turn depends on $(\alpha_F - \alpha_R)$ for a given steering angle and the wheel base l.

For steady-state cornering, α_F, α_R can directly be derived from the (27.63) through (27.65), with $\ddot{\psi} = 0$:

$$\alpha_F = \frac{l_R}{l C_F} m a_y, \quad \alpha_R = \frac{l_F}{l C_R} m a_y \tag{27.92}$$

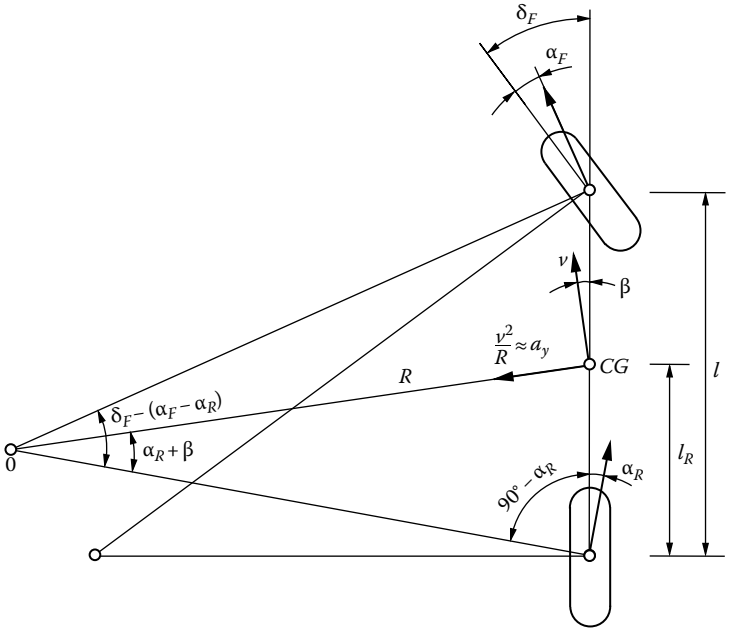


FIGURE 27.20 General steady-state cornering situation with center of rotation 0.

Comparing (27.90) with (27.87) results in

$$(\alpha_F - \alpha_R) = K_{US} a_{y,st} \tag{27.93}$$

and for $a_{y,st} > 0$, the relations and definitions for over/under/neutral steer in (27.89) are equally valid for $(\alpha_F - \alpha_R)$ instead of K_{US}.

The characterization of the steady-state behavior of a vehicle can now be seen in the plots of Figure 27.21. In Figure 27.21a, the driving velocity is kept constant and the increase in lateral acceleration is achieved by a change in the steering angle, while in Figure 27.21b the velocity is increased at constant radius R. For vehicle B with understeer characteristics, the steering angle has to be increased to keep radius R = constant with increasing velocity and has to be reduced for vehicle A with oversteer characteristics. In the latter case, the steering angle may even change sign at v_{crit}, (27.96), and countersteering will be necessary, if the velocity can further be increased. The required parameters for the two-wheel vehicle model of two different vehicles A and B are listed in Table 27.1.

In Figure 27.21c also, the side slip angles of the vehicles corresponding to Figure 27.21b are shown. Starting at $\beta = \beta_0$, (27.81) and (27.82), both curves $\beta = \beta\,(a_{y,st})$ change sign, and despite the different steering characteristics, the characteristics are in principal similar. Usually, the vehicle with oversteer characteristics shows larger absolute values of β with increasing $a_{y,st}$. Obviously, the side slip angle of the vehicle cannot be used for the characterization of the steering properties only.

With respect to a nonlinear system behavior, Figure 27.22 shows the steering characteristics and its common interpretation, for example, [6,7,17], for different types of vehicles, analogously to Figure 27.21a, for fixed velocity v_x. Gradient $d\delta_a/da_y = l/v_x^2$ decides between over- and understeer (see (27.86) and (27.89)). The condition for stability at fixed velocity v_x is $\partial\delta_F/\partial(l/R) > 0$ (see Section 27.3.3 for more details). Again, an unstable driving behavior may only occur for a vehicle with oversteer characteristics (see also Section 27.3.2.3).

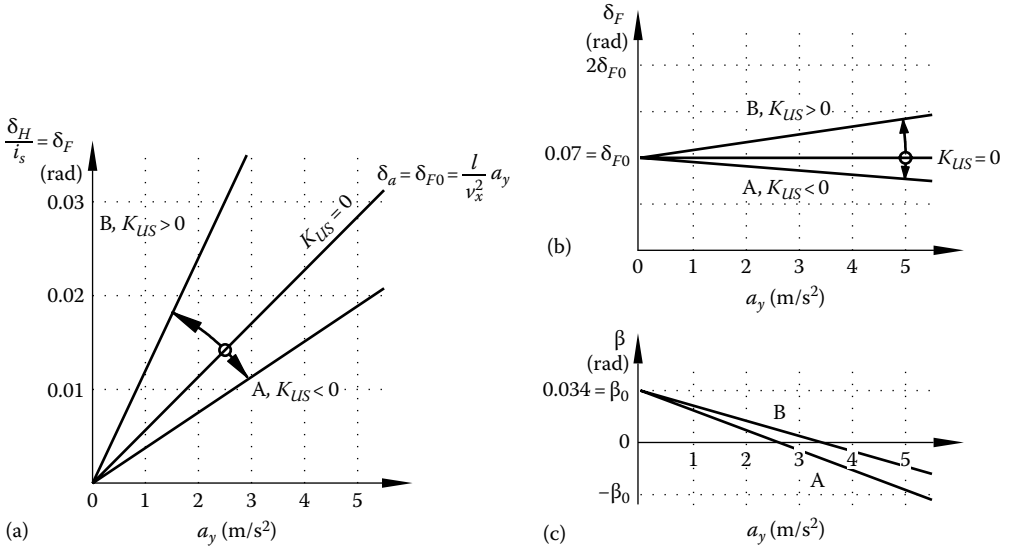


FIGURE 27.21 Steady-state steering characteristics, linear-handling diagram, vehicle parameters corresponding to Table 27.1: (a) for v = constant = 80 km/h; (b) for R = constant = 40 m; (c) side slip angles of the vehicle corresponding to (b).

TABLE 27.1
Vehicle Data for the Linear Two-Wheel Vehicle Model

Vehicle	A	B
m	1,900 kg	1,900 kg
I_Z	2,900 kg m²	2,900 kg m²
l_F	1.44 m	1.44 m
l_R	1.36 m	1.36 m
C_F	90,000 N/rad	60,000 N/rad
C_R	80,000 N/rad	110,000 N/rad
K_{US}	-1.96×10^{-3} rad s²/m	$+6.50 \times 10^{-3}$ rad s²/m
Steering characteristics	Oversteer	Understeer

Other characteristic quantities for steady-state cornering can be derived from transfer functions, like $G_r(s) = r(s)/\delta_F(s)$, $G_\beta(s) = \beta(s)/\delta_F(s)$, derived from Laplace transformation of (27.75) and (27.76). There results the steady-state (static) yaw velocity gain with respect to the input from steering hand wheel:

$$G_{r,st} = \left.\frac{r}{\delta_H / i_s}\right|_{st} = \frac{v_x}{l + K_{US} \cdot v_x^2} \tag{27.94}$$

The principally different behavior of vehicles with over- and understeer characteristics becomes apparent in Figure 27.23, where areas of stable or unstable system behavior are indicated as well (more details are provided later in this section).

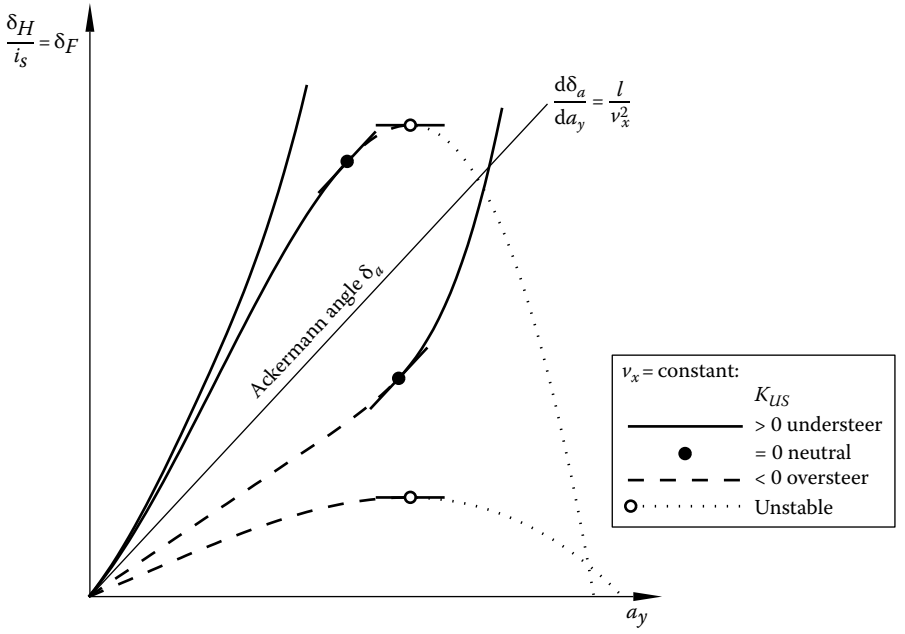


FIGURE 27.22 Principal nonlinear handling diagram, for v_x = constant, i_s = constant.

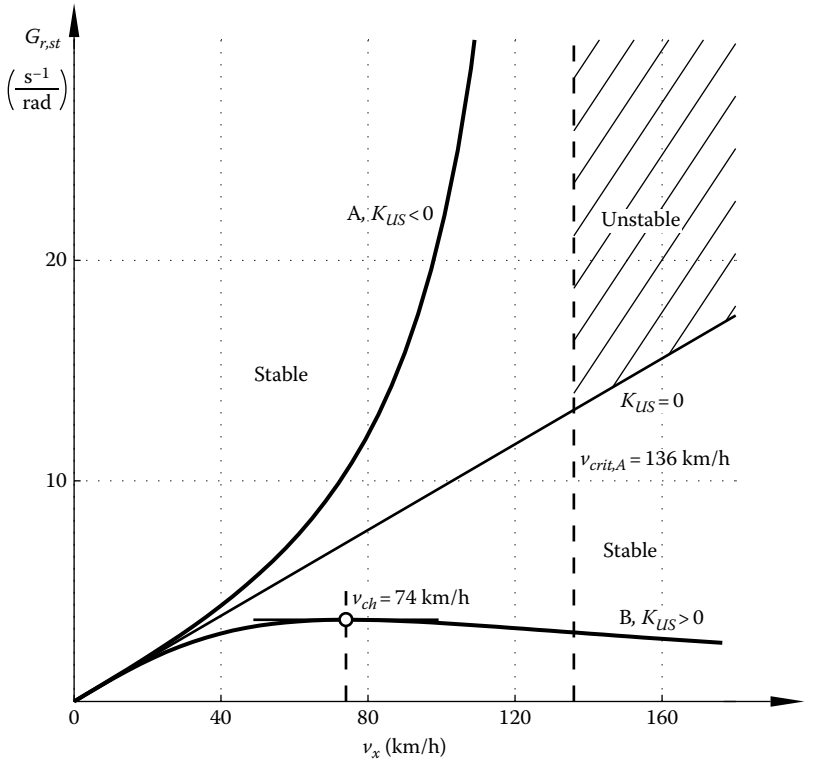


FIGURE 27.23 Behavior of vehicles with over-, neutral, and understeer characteristics with respect to the static yaw velocity gain; data for vehicles A and B according to Table 27.1.

For $K_{US} > 0$, vehicle B with understeer characteristics, the characteristic speed v_{ch} is defined by

$$\left.\frac{\partial G_r}{\partial v_x}\right|_{v_x = v_{ch}} = \frac{l - K_{US} v_{ch}^2}{(l + K_{US} v_{ch}^2)^2} := 0$$
$$\Rightarrow v_{ch}^2 = \frac{l}{K_{US}}, \quad K_{US} > 0 \tag{27.95}$$

and $G_{r,st}|_{v_x = v_{ch}} = v_{ch}/(2l)$ is half the one of a vehicle with neutral steer at $v_x = v_{ch}$.

For vehicle B with oversteer characteristics and $K_{US} < 0$, the static yaw velocity gain in (27.94) becomes infinity at the critical speed $v_x = v_{crit}$

$$G_{r,st}\big|_{v_x = v_{crit}} \to \infty$$
$$\Rightarrow v_{crit}^2 = -\frac{l}{K_{US}}, \quad K_{US} < 0 \tag{27.96}$$

The vehicle with neutral steer, $K_{US} = 0$, shows a steady increase in $G_{r,st}$ with velocity.

Consequently, one can find for the vehicles listed in Table 27.1:

$$v_{crit,A} = 37 \text{ m/s } (136 \text{ km/h}) \tag{27.97}$$

$$v_{ch,B} = 21 \text{ m/s } (74 \text{ km/h}) \tag{27.98}$$

In general, modern vehicles do not show oversteer characteristics for a wide range of conditions (and understeer behavior for at least the linear region of steering characteristics); thus, it is essential to choose a proper characteristic velocity v_{ch}. At this speed, the vehicle is most sensitive to steering inputs w.r.t. yaw motion. In [22], a range between 65 and 100 km/h is therefore recommended.

As a matter of fact, even with this kind of simple vehicle model, first conclusions can be drawn of how to increase or improve understeer characteristics of a car from a practical point of view. A quick look at the definition of the understeer gradient K_{US} in (27.88) reveals the constitutive term $(C_R l_R - C_F l_F)$ thereof, and consequently, $C_R l_R$ has to be increased and/or $C_F l_F$ decreased. Keeping in mind that the effective cornering stiffness comprises both tire and suspension behavior, the (elasto) kinematic design of the front and rear suspension and different tire characteristics (dimension, inflation pressure, etc.) are effective ways to influence the steering and handling characteristics by C_F,C_R. The application of traction forces will reduce the cornering stiffness; thus, for example, a trend toward understeer can be attained for a vehicle with front-wheel drive. Putting more load on the rear wheel(s) will decrease l_R and increase C_R considering nonlinear tire characteristics for this purpose. Due to the nonlinear relation between C_i $(i = F, R)$ and the tire load, the increase of C_R is less than the decrease of C_F and K_{US} will be reduced toward more oversteer. A similar explanation holds for a change in the tire load distribution at one axle. The force distribution at the front axle can be increased by applying stiffer spring characteristics of the suspension and/or a torsion bar and thus C_F of the substitutive front wheel can be reduced. Although the understeer gradient is raised in this way, the stiffer suspension may influence ride comfort adversely. For details on the effects of suspension design, see Chapter 21.

The steady-state (static) lateral acceleration gain $G_{ay,st}$ directly follows from (27.94)

$$a_{y,st} = r_{st} \cdot v_x$$
$$G_{ay,st} = \left.\frac{a_y}{\delta_H / i_s}\right|_{st} = G_{r,st} \cdot v_x = \frac{v_x^2}{l + K_{US} \cdot v_x^2} \tag{27.99}$$

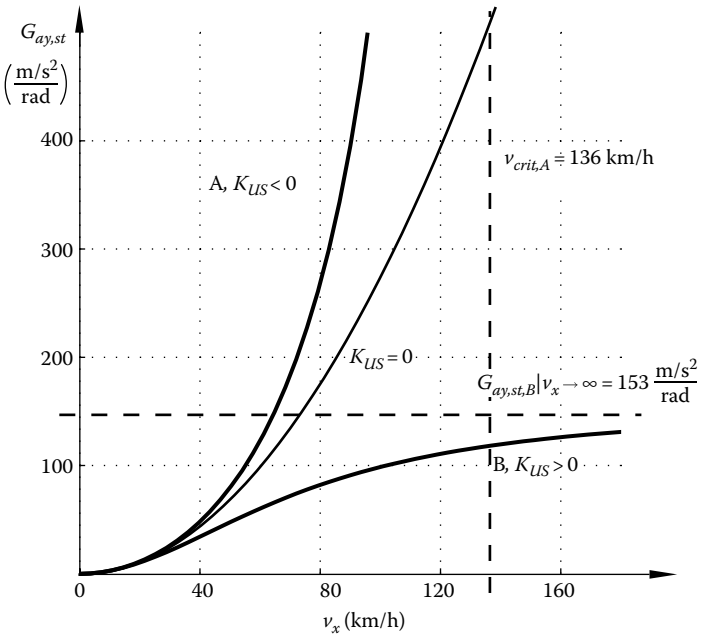


FIGURE 27.24 Behavior of vehicles with over-, neutral, and understeer characteristics with respect to the static lateral acceleration gain; data for vehicles A and B according to Table 27.1.

and consequently $G_{ay,st} \to \infty$ at the same velocity v_{crit} for $K_{US} < 0$, vehicle A in Figure 27.24. For vehicle B with understeer characteristics, $K_{US} > 0$, no equivalent of v_{ch} can be found; here, the gain $G_{ay,st}$ converges to a finite value for large velocities. For neutral steer, $K_{US} = 0$, the gain increases to ∞ with $v_x \to \infty$.

Considering the steady-state gains $G_{r,st}$ and $G_{ay,st}$, Figures 27.23 and 27.24, it becomes obvious that a vehicle with understeer characteristics has a more moderate response compared to the vehicle with oversteer. The steady-state (static) side slip gain $G_{\beta,st}$ of the vehicle may easily be taken from (27.75).

27.3.2.3 Stability and Nonsteady-State Vehicle Behavior

Based on (27.72) through (27.74) or (27.75) through (27.78), the stability behavior of the vehicle, the vehicle response to harmonic steering input (Bode diagrams), and a steering step input will be discussed. The yaw velocity transfer function to steering input $G_r(s)$ is a common way to characterize nonsteady-state vehicle behavior [7,22]. Considering (27.76) and taking into account (27.94) yields

$$\begin{aligned} G_r(s) = \frac{r(s)}{\delta_F(s)} &= G_{r,st} \frac{1 + T_Z s}{\omega_0^2 + 2D\omega_0 s + s^2} \cdot \omega_0^2 \\ &= G_{r,st} \frac{1 + T_Z s}{1 + (2D/\omega_0)s + \left(s^2/\omega_0^2\right)} \end{aligned} \tag{27.100}$$

with

$$\begin{aligned} T_Z &= \frac{m v_x l_F}{C_R l} \\ \omega_0^2 = K_2 &= \frac{C_F C_R l^2}{I_Z m v_x^2} \left(1 + \frac{K_{US} v_x^2}{l}\right) \\ D\omega_0 &= K_1 \end{aligned} \tag{27.101}$$

K_1, K_2 from (27.77), (27.78). The eigenfrequency ω of the damped system is

$$\omega^2 = \omega_0^2(1-D^2) \tag{27.102}$$

By applying (27.84) and (27.87) in (27.101), it can be concluded directly from steady-state steering properties to the eigenfrequency ω_0 of the undamped system:

$$\omega_0^2 = \frac{C_F C_R l^2}{I_Z m v_x^2}\left(\frac{\delta_{F,st}}{\delta_a}\right) = K_2 \tag{27.103}$$

Stability of the vehicle behavior is determined by the equal homogeneous parts of the differential equations (27.75) and (27.76) or by the system matrix $\mathbf{F}$ in (27.73) from (27.72); the eigenvalues $\lambda_{1,2}$

$$\lambda_{1,2} = -K_1 \pm \sqrt{K_1^2 - K_2} \tag{27.104}$$

can for instance easily be derived from setting the denominator in (27.100) to zero or from $|\mathbf{F}-\lambda\mathbf{I}|=0$, that is the characteristic equation. Since always $K_1 > 0$, according to (27.77), the sign of $K_2 = \omega_0^2 = K_2\left(v_x^2\right)$ determines the stability of the motion of the system with respect to small disturbances in the initial conditions. This directly follows from the Hurwitz stability criterion for a second-order differential equation, $K_1 > 0$, $K_2 > 0$. Essentially, stability means that the system trajectory can be kept arbitrarily close to the equilibrium state by starting sufficiently close to it ("Lyapunov stability").

The eigenfrequency $\omega_{0|u}$ of the undamped system for a vehicle with understeer characteristics, with $K_{US} > 0$ and $v_{ch}^2 = l/K_{US} > 0$, (27.95), is

$$\omega_{0|u}^2 = K_2 = \frac{C_F C_R l^2}{I_Z m v_x^2}\left(1 + \frac{v_x^2}{v_{ch}^2}\right) > 0 \tag{27.105}$$

and the steady-state motion of the vehicle is stable for all driving velocities ($K_2 > K_1^2$ resulting in a negative complex pair of eigenvalues, $K_2 < K_1^2$ in two negative real eigenvalues).

For a vehicle with oversteer characteristics, $K_{US} < 0$, and

$$\begin{aligned} v_{crit}^2 &= -\frac{l}{K_{US}} \\ \omega_{0|0}^2 = K_2 &= \frac{C_F C_R l^2}{I_Z m v_x^2}\left(1 - \frac{v_x^2}{v_{crit}^2}\right) \begin{matrix} \leq \\ \geq \end{matrix} 0 \end{aligned} \tag{27.106}$$

it becomes obvious that $\omega_{0|0}^2 < 0$, for velocities $v_x > v_{crit}$, and the steady-state motion of the vehicle will then become unstable with one positive and one negative real eigenvalue. Please notice that the steady-state motion of a vehicle with oversteer characteristics is not unstable a priori, but there is a stability margin at a certain speed v_{crit}!

In Figure 27.25, natural damping factor D and eigenfrequencies ω_0, ω are plotted for the two vehicles A and B, Table 27.1, as a function of the driving velocity v_x. Due to $D > 1$ for vehicle A with oversteer characteristics, there does not exist an eigenfrequency of the damped system—the system has aperiodic solutions. For a typical vehicle with understeer characteristics, vehicle B, the eigenfrequency of the damped system starts from infinity to a finite value for $v_x \to \infty$; the natural damping factor $D < 1$ decreases with increasing velocity. Therefore, a driver may handle oscillations of the yaw rate, induced, for example, by an evasive maneuver on a highway, with difficulties, as the vehicle motion is less damped at high speed, as might be expected by the driver; also, eigenfrequencies vary with speed.

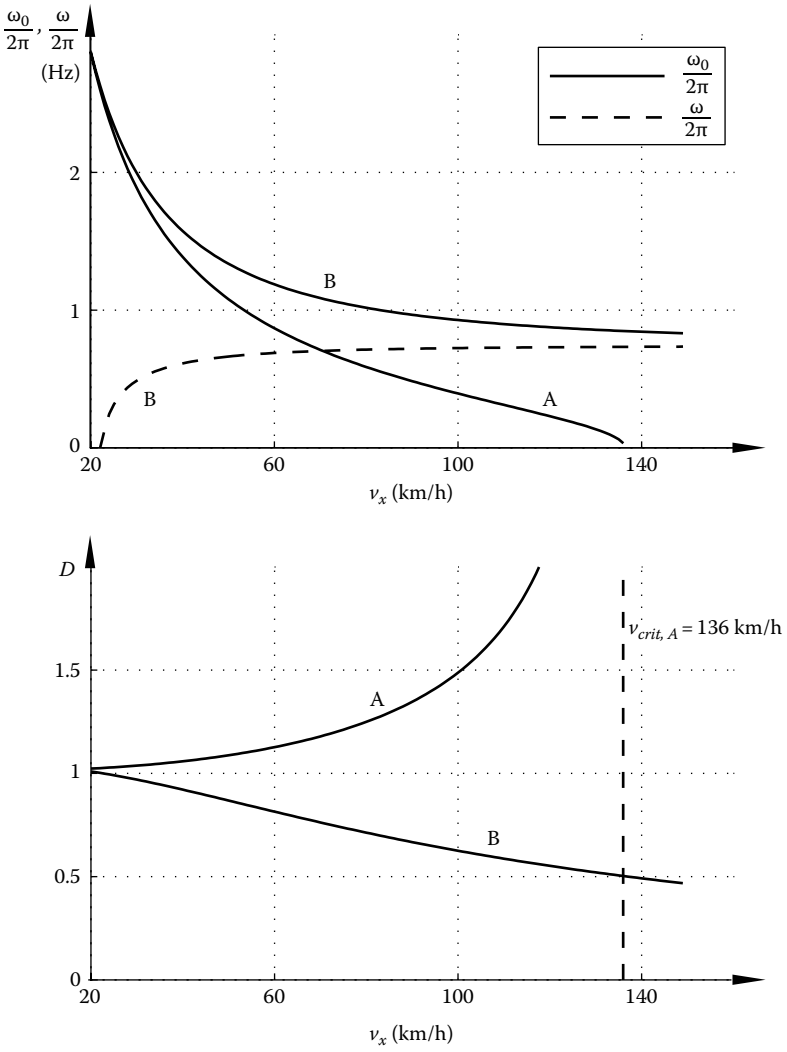


FIGURE 27.25 Eigenfrequency $f_0 = \omega_0/2\pi$, eigenfrequency $f = \omega/2\pi$ of the damped system, and damping factor D as a function of velocity v_x for vehicles A and B (data according to Table 27.1).

The response of the vehicle to a *steering step input*, respectively a very quick application of a specific steering angle (*steering ramp*) at real test conditions, is often used to analyze its sensitivity with respect to handling maneuvers. It is characterized by parameters illustrated principally in Figure 27.26, according to test procedures of road vehicles in ISO 7401. An important quantity therein is the peak-response-time T_{rmax} (typically 200–400 ms in modern passenger cars, [22]), as it shows a good correlation to the subjective evaluation of the steering step/ramp response behavior.

The plotted time history is typical for a damped system of second order, (27.76), with a negative complex pair of eigenvalues, $K_2 > K_1^2$ in (27.104), and corresponds therefore to a vehicle with understeer characteristics. The transient response for a steering step input, for example, [7,22], with D, ω, ω_0, T_Z from (27.101), is then

$$\frac{r(t)}{\delta_{H,st}/i_s} = G_{r,st}\left[1 - \mathrm{e}^{-D\omega_0 t}\left(\cos\omega t + \frac{D - T_Z\omega_0}{\sqrt{1-D^2}}\sin\omega t\right)\right] \tag{27.107}$$

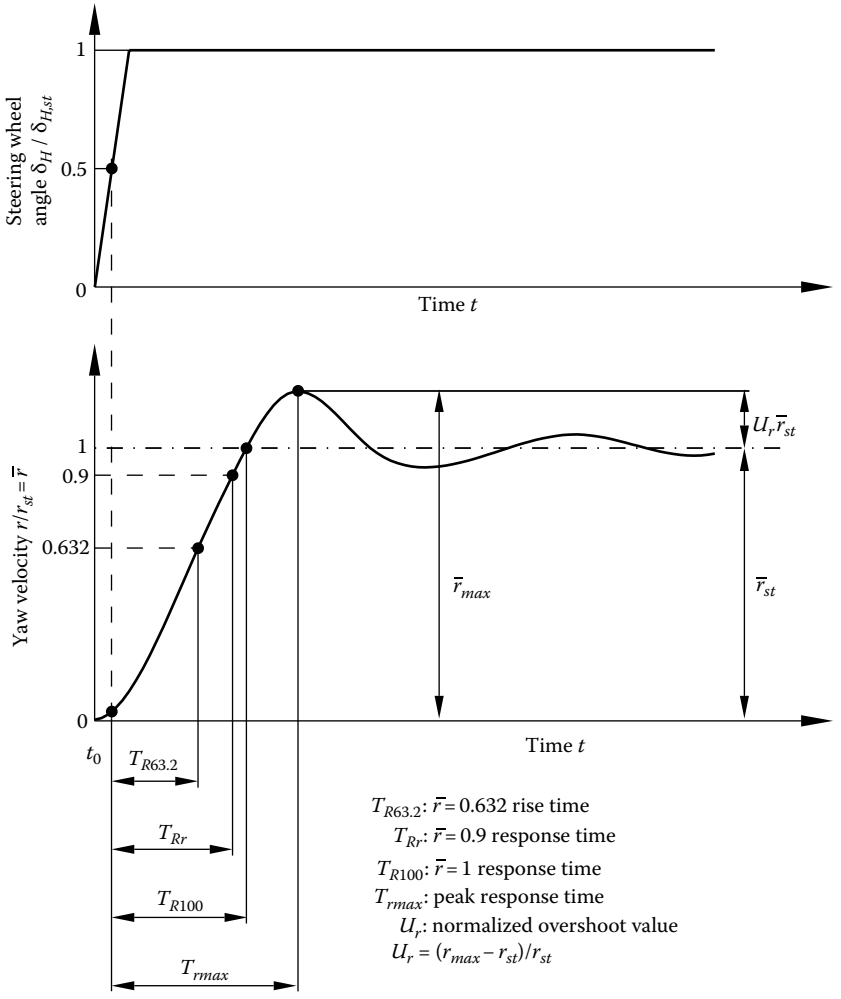


FIGURE 27.26 Transient yaw motion characteristics after steering ramp.

The initial gradient is determined by $\dot{r}(0)/(\delta_{H,st}/i_s) = (C_F l_F)/I_Z$, directly from (27.74) in (27.72) with $v_y(0) = r(0) = \delta_R = 0$, for vehicles with understeer or oversteer characteristics. For both characteristics and $0 < K_2 > K_1^2$ in (27.104), the two negative real eigenvalues result in an aperiodic yaw motion to $G_{r,st}$ for $t \to \infty$. According to Figure 27.23, $G_{r,st}$ is substantially larger and therefore the radius of the finally resulting steady-state cornering is smaller for a vehicle with oversteer than with understeer characteristics for the same steering input at the same speed. If the speed of the vehicle with oversteer is beyond v_{crit} and a steering step/ramp is applied, the vehicle will turn-in with exponentially increasing yaw rate, and the driver has to interact appropriately.

For the yaw rate response, a limited area was proposed by Experimental Safety Vehicle (ESV) (see [12] and Figure 27.27 for a van). The two plotted curves of experimental tests indicate where critical points in the area may occur. The origin at $t = 0$ should coincide with t_0 of Figure 27.26.

Active body control, additional rear wheel steering, and other active control systems can be designed to improve the handling and stability behavior and may enhance the response of the

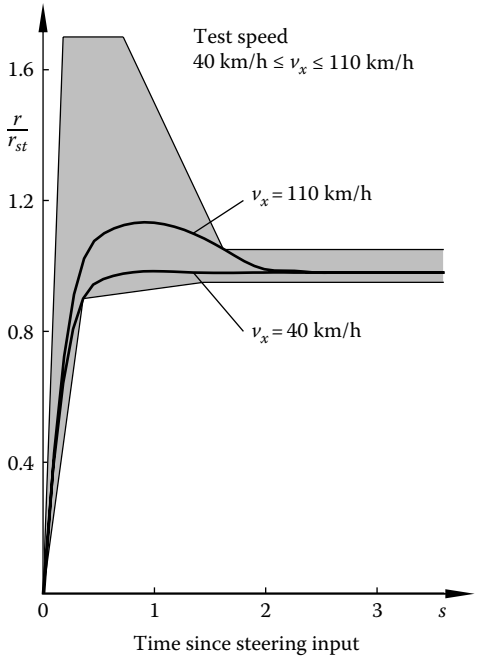


FIGURE 27.27 ESV limits for a steering step/ramp input ($\dot{\delta}_H \approx 500°/s$) with final steady-state lateral acceleration ($a_{y,st} \approx 0.4\ g$).

vehicle with respect to the input of the driver [23]. Then the driver may need to adapt himself to the changed steering behavior as well (Figure 27.27).

The dynamic behavior can also be analyzed with *frequency responses* (*Bode diagrams*). It is an alternative, if at very high speeds, steering step inputs cannot be performed. The Bode diagram of the yaw velocity with respect to harmonic steering input can be derived directly from the transfer function (27.100) and by replacing s with $i\nu$ (ν is excitation frequency, $i^2 = -1$),

$$\left.\frac{r(s)}{\delta_F(s)}\right|_{s=i\nu} = G_r(i\nu) = G_{r,st}\,\frac{1+T_Z(i\nu)}{1+\left(2D/\omega_0\right)(i\nu)+(i\nu)^2/\omega_0^2} \tag{27.108}$$

Therefrom, the magnitude $A(\nu) = |G_r(i\nu)|$ or $A(\nu) = 20\ \log\ |G_r(i\nu)|$ in dB and phase diagram $\phi(\nu) = \arctan[\Re(G_r(i\nu))/\Im(G_r(i\nu))]$ can be plotted.

Measured frequency responses of a vehicle with understeer and K_{US} very close to vehicle B are shown in Figure 27.28 [22]. The frequencies start from 0.2 to about 3 Hz, the feasible range for a human driver.

On the left side of Figure 27.28, the Bode diagram of the yaw rate is shown. At $\nu = 0$, the magnitude is, in theory, equal to $G_{r,st} = G_{r,st}(v_x)$; its increase at moderate speeds, here from $v_x = 60$–80 km/h, becomes obvious, and for higher speeds above the characteristic velocity at $v_{ch} = 76$ km/h, not much influence on the velocity can be noticed (compare with Figure 27.23). The theoretically horizontal tangent to $A(\nu)$ at $\nu = 0$ is indicated by a small increase in magnitude for $\nu \to 0$ in the measured plot. The behavior of magnitude and amplification of resonance around the eigenfrequency depend notably on the natural damping factor D, which are both functions of v_x, compare with Figure 27.25, and what is clearly reflected in the figure. The phase angle and the considerable delay of the yaw rate with respect to the steering angle input for higher frequencies is illustrated in the bottom left plot.

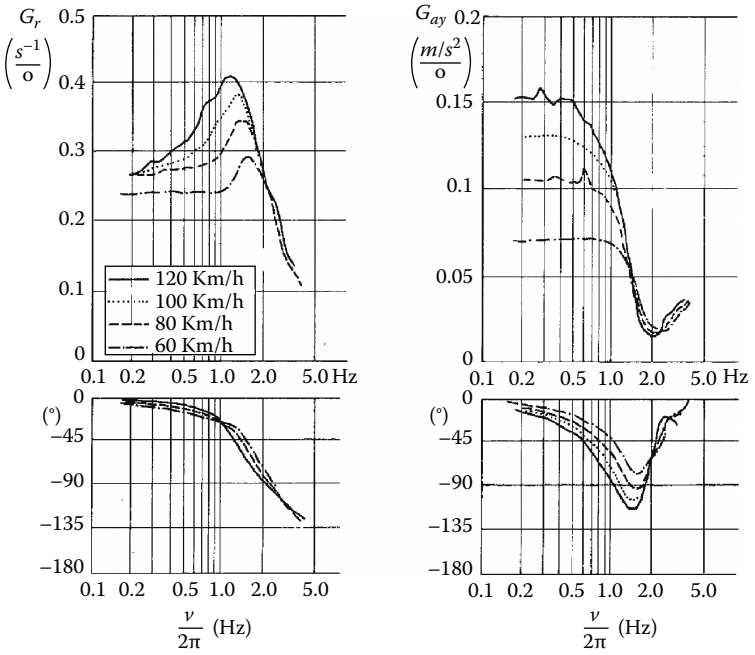


FIGURE 27.28 Measured yaw velocity and lateral acceleration frequency responses of a vehicle with understeer characteristics ($K_{US} = 0.00620$ rad s^2/m, $v_{ch} = 76$ km/h) for different driving velocities v_x. (From Zomotor, A., *Fahrwerktechnik: Fahrverhalten*, Reimpell, J. (ed.), Vogel Buchverlag, Würzburg, Germany, 1987.)

The frequency response of the side slip angle of the vehicle may be derived from (27.75) without much effort, but in general the frequency response of the lateral acceleration is of more interest. The three frequency responses are related to each other by their transfer functions (see (27.70)),

$$\frac{a_y(s)}{\delta_F(s)} = v_x\left[\frac{r(s)}{\delta_F(s)} + \frac{s\beta(s)}{\delta_F(s)}\right]$$

There follows the frequency response of the lateral acceleration

$$\left.\frac{a_y(s)}{\delta_F(s)}\right|_{s=i\nu} = G_{ay}(i\nu) = G_{ay,st}\,\frac{1+T_1(i\nu)+T_2(i\nu)^2}{1+(2D/\omega_0)(i\nu)+(i\nu)^2/\omega_0^2} \tag{27.109}$$

with

$$T_1 = \frac{l_R}{v_x}, \quad T_2 = \frac{I_Z}{C_R l}$$

The measured Bode diagram of the lateral acceleration is presented on the right side of Figure 27.28. The magnitudes start at $\nu \to 0$ at very different levels of $G_{ay,st} = G_{ay,st}(v_x)$, see also Figure 27.24, keep more or less constant, before a very strong decline can be observed, without pronouncing eigenfrequencies. On the one hand, the drop in the magnitude may not be at too low frequencies to keep the vehicle sensitive for the driver's steering input; on the other hand, the constant magnitude in the range of moderate steering frequencies at moderate velocities accounts for an easier controllability. The influence of velocity on the phase delay of the lateral acceleration is more distinct than that of the yaw rate.

Finally, the Bode diagram of the $G_{ay,st}$-normalized lateral acceleration for vehicle A and B is shown in Figure 27.29. The characteristics of vehicle B with understeer are very similar to the

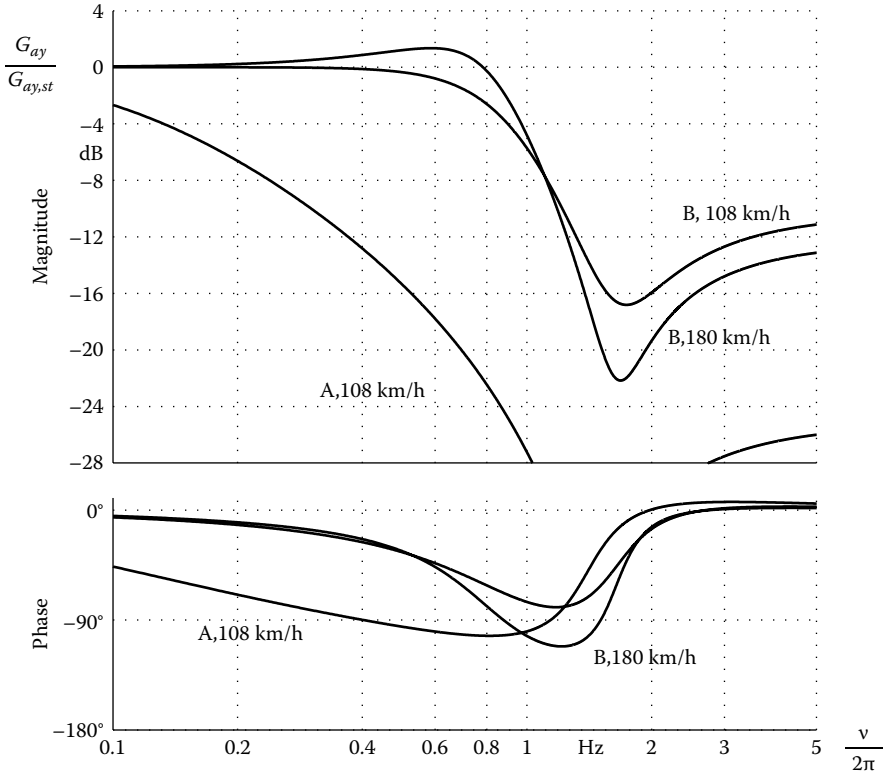


FIGURE 27.29 Normalized lateral acceleration frequency response of vehicle A with oversteer and vehicle B with understeer characteristics (see Table 27.1).

previous one discussed in Figure 27.28. Only for very high speeds, 180 km/h, when the damping coefficient becomes small according to Figure 27.25, a resonance peak can be noticed at the magnitudes corresponding to the eigenfrequency in Figure 27.25. Since the velocity of 108 km/h is already close to the critical speed v_{crit} = 136 km/h of vehicle A with oversteer, (27.97), the response of this vehicle in the area of convenient, normal steering until about 1 Hz very much depends on the frequency (exhibited by a strong decline) and reveals an increased phase lag. Since 180 km/h is above the critical speed of vehicle A, the corresponding characteristics are omitted for this vehicle. Vehicle B with understeer characteristics provides a much more driver-friendly behavior with small changes in magnitude and smaller phase delay angles, even for a velocity of 180 km/h. For $\nu > 1$ Hz, even vehicle B with understeer characteristics displays major changes in magnitude.

27.3.3 Nonlinear Two-Wheel Vehicle Model

The assumptions for the application of a linear two-wheel vehicle model hold for moderate lateral accelerations up to about 4 m/s^2 on a high friction surface and constant velocity (or small longitudinal accelerations). Exceeding these limits calls for more complex vehicle models, in particular, a four-wheel vehicle model. Nowadays, vehicle models on the basis of multibody-system formulation and programs are often employed (see Chapter 2) rather than simplified four-wheel vehicle models with a symbolic representation of the equations of motion (e.g., [7,8]). The complex vehicle model allows for a detailed representation of all essential subsystems of the vehicle to analyze and simulate subsystem and global vehicle system dynamics efficiently; therefore, please refer to Chapter 28.

Between these ends, the nonlinear two-wheel vehicle model is an intermediate stage, providing still enough simplicity to understand basic (nonlinear) phenomena without neglecting the most crucial nonlinearities, which are—most often—nonlinear tire characteristics. Besides linearized tire

characteristics, geometrical linearization of trigonometric functions, for example, $\sin\delta \approx \delta$, $\cos\delta \approx 1$, and $\tan\delta \approx \delta$, facilitates an easier interpretation of the system equations. While the later linearization holds up to angles of about 10°–15°, the former applies for side slip angles up to about 5° only, necessitating a nonlinear tire model beyond. In the following, these two simplifications are partially or fully abandoned to derive and evaluate the *nonlinear steady-state handling diagram* still on the basis of a two-wheel vehicle model.

The nonlinear steady-state equations of the vehicle model in Figure 27.18 are ($M_z = W_Y = \delta_R = 0$, $a_x \approx 0$)

$$-m\frac{v^2}{R}\sin\beta = F_{xF}\cos\delta_F - F_{yF}\sin\delta_F + F_{xR} - W_L \tag{27.110}$$

$$m\frac{v^2}{R}\cos\beta = F_{xF}\sin\delta_F + F_{yF}\cos\delta_F + F_{yR} \tag{27.111}$$

$$I_Z\ddot{\psi} = 0 = (F_{xF}\sin\delta_F + F_{yF}\cos\delta_F)l_F - F_{yR}l_R \tag{27.112}$$

The influence of the side slip angle of the vehicle on the aerodynamic drag coefficient in W_L is usually neglected. Forces $F_{xi}(i = F, R)$ are either rolling resistance or traction force (of a front- or rear-wheel drive) for constant velocity v at steady-state cornering with constant radius R. The lateral axle forces $F_{yi}(i = F, R)$ depend on the respective side slip angles, on the vertical axle loads $F_{zi}(i = F, R)$, and on the longitudinal tire forces, but the latter are often not taken into account:

$$F_{yF} = F_{yF}(\alpha_F, F_{zF}) \tag{27.113}$$

$$F_{yR} = F_{yR}(\alpha_R, F_{zR}) \tag{27.114}$$

$$F_{zF} = mg\frac{l_R}{l}, \quad F_{zR} = mg\frac{l_F}{l} \tag{27.115}$$

These characteristics can for instance be described by the magic formula [8]

$$F_y = D\sin\{C\arctan[B\alpha - E(B\alpha - \arctan(B\alpha))]\} \tag{27.116}$$

with parameters μ, c_α, C, E, the peak lateral substitutive tire/axle force $D = \mu F_z$, and the corresponding stiffness $C_\alpha = BCD = c_\alpha F_z$; therefore, $B = c_\alpha/(C\mu)$. Other tire models are to be found in [24].

The nonlinear relations for (27.90) and (27.91) in Figure 27.20 are as follows:

$$\delta_F = \arctan\left[\frac{l_F + R\sin\beta}{R\cos\beta}\right] - \alpha_F \tag{27.117}$$

$$\beta = \arcsin\left[\frac{l_R\cos\alpha_R}{R}\right] - \alpha_R \tag{27.118}$$

For given v, the seven equations (27.110) through (27.114), (27.117), and (27.118) can now be numerically solved for the zeros: β, α_F, α_R, δ_F, or R (depending on problem), F_{yF}, F_{yR}, F_{xF}, or F_{xR} (for front- or rear-wheel drive). Alternatively, an iterative solution scheme can be applied, which is very

effective and usually converges quickly [19]. Therefore, (27.110) through (27.112) are rewritten for a vehicle with front-wheel drive

$$\begin{bmatrix} \cos\delta_F & -\sin\delta_F & 0 \\ \sin\delta_F & \cos\delta_F & 1 \\ l_F\sin\delta_F & l_F\cos\delta_F & -l_R \end{bmatrix} \begin{bmatrix} F_{xF} \\ F_{yF} \\ F_{yR} \end{bmatrix} = \begin{bmatrix} -m\dfrac{v^2}{R}\sin\beta - F_{xR} + W_L \\ m\dfrac{v^2}{R}\cos\beta \\ 0 \end{bmatrix} \tag{27.119}$$

and there follows

$$\begin{aligned} F_{xF} &= m\frac{v^2}{Rl}[l_R\sin\delta_F\cos\beta - l\cos\delta_F\sin\beta] - (F_{xR} - W_L)\cos\delta_F \\ F_{yF} &= m\frac{v^2}{Rl}[l_R\cos\delta_F\cos\beta + l\sin\delta_F\sin\beta] + (F_{xR} - W_L)\sin\delta_F \\ F_{yR} &= m\frac{v^2}{Rl}l_F\cos\beta \end{aligned} \tag{27.120}$$

Analogously for rear-wheel drive,

$$\begin{aligned} F_{xR} &= m\frac{v^2}{Rl\cos\delta_F}[l_R\sin\delta_F\cos\beta - l\cos\delta_F\sin\beta] + W_L - \frac{F_{xF}}{\cos\delta_F} \\ F_{yF} &= m\frac{v^2 l_R\cos\beta}{Rl\cos\delta_F} - F_{xF}\tan\delta_F \\ F_{yR} &= m\frac{v^2}{Rl}l_F\cos\beta \end{aligned} \tag{27.121}$$

Iteration starts by applying $\delta_{F0} = l/R$ and $\beta_0 = l_R/R$ for low-speed steering in (27.120) or (27.121) to derive the forces. The side slip angles α_F, α_R of the axles are obtained from the tire characteristics, (27.113) and (27.114), where the interaction between longitudinal and lateral tire forces may also be taken into account. Finally, (27.117) and (27.118) yield new starting values δ_F and β for the next iteration step. In this way, the nonlinear handling diagram according to Figure 27.22 is computed.

A very illustrative method to construct the nonlinear handling diagram, see [8], can be applied, when angles are small and the linear relation (27.90) still holds, and further $F_{xF}\sin\delta_F$ can be neglected in (27.111) and (27.112). Nevertheless, either a nonlinear mathematical representation of the lateral axle characteristics, for example, (27.116), or measured axle characteristics including suspension, steering, and other effects can be involved.

The steady-state lateral axle forces are derived from (27.63) and (27.64) with $\ddot{\psi} = 0$ to $F_{yF} = (ma_y l_R)/l$, $F_{yR} = (ma_y l_F)/l$, and applying (27.115) yields

$$\frac{F_{yF}}{F_{zF}} = \frac{F_{yR}}{F_{zR}} = \frac{a_y}{g} = \frac{v^2}{Rg} \tag{27.122}$$

These ratios can also be plotted using numerical tire characteristics, (27.113) through (27.115), respectively (27.116), or measured axle characteristics as a function of the side slip angle (see top plot

in Figure 27.30). Subtracting these characteristics horizontally from each other gives $\alpha_F - \alpha_R$ for the respective normal acceleration $v^2/(Rg)$ according to (27.122), left part of the bottom plot of the figure. For fixed $v^2/(Rg)$ and wheel track l, the diagram is extended to the right by plotting various fixed velocity v lines relating to $l/R = \delta_a$ (to quantify the radius R), and the corresponding steering angle can finally be read off from the sum $\delta_{F,v} = l/R + (\alpha_F - \alpha_R)$, (27.90), that is the distance between the handling curve and the respective straight velocity line. In other words, for a point on a given constant velocity line, a_y/g is found on the vertical axis, l/R on the right horizontal axis, and δ_F on the horizontal line cut by the handling curve. The nonlinear handling diagram is completed by running through various lateral accelerations (Figure 27.30), with points on the lower half plane belonging to negative R, a right-hand turn. The handling diagram presented corresponds to vehicle A (with oversteer) in Table

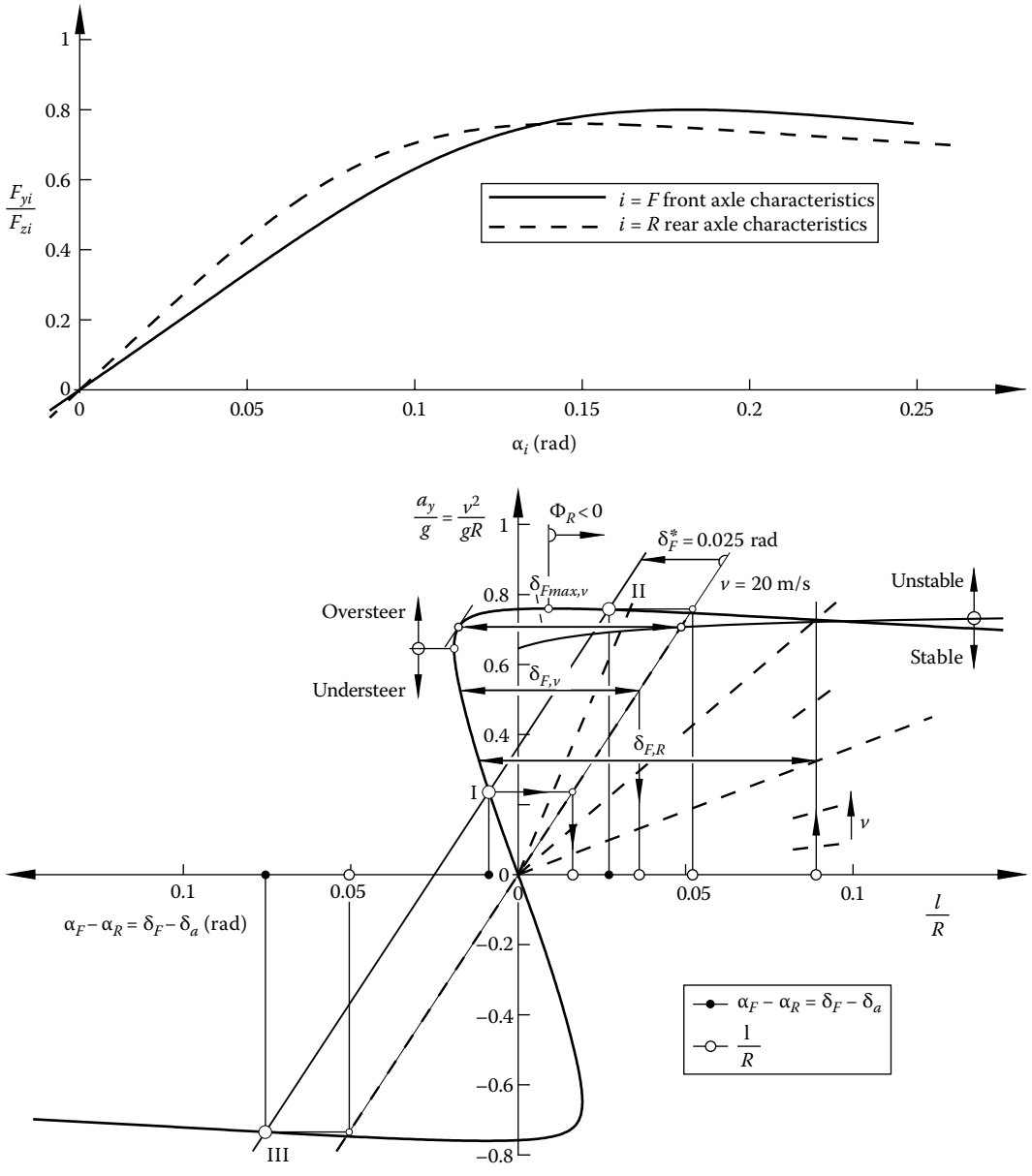


FIGURE 27.30 Nonlinear handling diagram.

27.1, and the (changed) normalized tire characteristics are plotted on top of Figure 27.30. The lower maximum of the two curves defines the maximum possible lateral acceleration a_y/g in the handling diagram.

If the radius R of the turn is kept constant, l/R is given; the required steering angle $\delta_{F,R}$ follows from the distance between the handling curve and the straight vertical line at l/R at the intersection with the selected velocity (line).

The definition of the understeer gradient in (27.86) following SAE J670e [17] does not explicitly mention the test conditions, which has no effect in the linear case. But for nonlinear tire characteristics [25], it makes a difference if v or $1/R$ is considered as an independent variable:

$$\left.\frac{\mathrm{d}(\delta_F - l/R)}{\mathrm{d}a_y}\right|_{v=\text{const.}} = \frac{gl}{v^2}\left[\frac{\partial \delta_F}{\partial (l/R)} - 1\right] \tag{27.123}$$

$$\left.\frac{\mathrm{d}(\delta_F - l/R)}{\mathrm{d}a_y}\right|_{R=\text{const.}} = \frac{gR}{2v}\frac{\partial \delta_F}{\partial v} \tag{27.124}$$

The condition $\mathrm{d}(\delta_F - l/R)/\mathrm{d}a_y = 0$ for neutral (and related over-/understeer) steer characteristics yields $\partial\delta_F/\partial(l/R)|_{v\,=\,\text{const.}} = 1$ from (27.123), and $\partial\delta_F/\partial v|_{R\,=\,\text{const.}} = 0$ from (27.124), which may result in different operating points. While Figure 27.22 is based on (27.123), Figure 27.30 refers to (27.124) and consequently defines a tangent to the handling curve with a slope to the left as understeer ($\partial\delta_F/\partial v > 0$) and to the right as oversteer characteristics ($\partial\delta_F/\partial v < 0$); the gradient at the origin corresponds to the linear(ized) system with the understeer gradient K_{US} in (27.93). Thus, the shape of the tire characteristics or rather the axle characteristics, which include effects from the suspensions, traction force, etc., determine handling and stability characteristics of the vehicle. If the shapes show a maximum with a downward slope at large side slip angles, the handling curve becomes more complex, and apart from the main branch through the origin, other (isolated) branches may occur (see [8]). Here, the normalized tire characteristics are cut early enough for reasons of simplification.

For stability analysis [8], the nonlinear equations of motion, (27.110) through (27.112), as well as the normalized nonlinear tire characteristics are linearized, $\Phi_i = (1/F_{zi})(\partial F_{yi}/\partial\alpha_i)|_{st}$ ($i = F, R$), with respect to steady-state solutions ($|_{st}$). The Hurwitz criterion, necessary and sufficient, then yields two conditions for stability (compare to (27.77) and (27.78)):

$$(I_Z + ml_F^2)l_R\Phi_F + (I_Z + ml_R^2)l_F\Phi_R > 0 \tag{27.125}$$

$$\Phi_F\Phi_R\left.\frac{\partial \delta_F}{\partial (l/R)}\right|_{v=\text{const.}} > 0 \tag{27.126}$$

The first condition is usually of minor importance: the absolute value of a negative slope Φ_i is in most cases much smaller than the positive slope. Since Φ_F and Φ_R are of the same sign in virtually all practical operating conditions, they are often dropped in the second condition (27.126). That implies for Figure 27.22, instability can only occur, when the vehicle shows oversteer characteristics and $\partial\delta_F/\partial(l/R)|_{v\,=\,\text{const.}} < 0$. In the handling diagram of Figure 27.30, the stability boundary can be constructed pointwise according to (27.126) by finding the tangent to the handling curve that runs parallel to the velocity (line) considered. Stability is maintained as long as δ_F increases with increasing l/R for constant velocity; beyond the maximum $\delta_{Fmax,v}$, the motion becomes unstable. The handling curve beyond its maximum, wherefrom $\Phi_R < 0$, is still associated with unstable motions, as $\partial\delta_F/\partial(l/R)|_{v\,=\,\text{const.}} > 0$ and $\Phi_F > 0$.

For a predefined velocity ($v = 20$ m/s) and steering angle ($\delta_F^* = 0.025$ rad), corresponding radii to steady-state cornering can be picked from the handling diagram (Figure 27.30) without much effort. The respective velocity line is shifted in parallel over a distance corresponding to δ_F^* to the left. The three resulting intersection points I, II, and III mark the three possible points of equilibrium; the related cornering radii, l/R, are found by returning to the velocity line and to the abscissa. Condition III belongs to a right-hand curve with negative cornering radius. While the motion in I is stable, it is unstable for II and III, considering the curve of the boundary for stability in the figure; more on stability is revealed in the following. The intersection points of the velocity lines through the origin with the handling curve represent steady-state solutions with zero δ_F, the one in the origin refers to a (stable) straight ahead motion.

Numerical integration of the system equations (27.63) and (27.64) with (27.66) and respective (nonlinear) axle characteristics, (27.113) through (27.115) with (27.68), yields for a given constant velocity and steering angle $\delta_F = \delta_F(t)$, the state variables $\dot{\psi} = r = r(t)$, and $v_y = v_y(t)$, and by further integration the trajectory of CG and the yaw angle $\psi = \psi(t)$ of the vehicle. Also, stability may be addressed in that way. Phase-plane analysis of a nonlinear second-order system is a graphical method to draw conclusions regarding stability after large disturbances without the complete integration of the system. Also, the nature (node, focus, saddle, limit cycle) of the equilibrium points (singular points) can be determined, boundaries of the domain of attraction may be found, if they exist, and therefrom concluded to local stability (with stability boundaries) or global stability in certain cases.

For that purpose, time is eliminated from (27.63) and (27.64) with (27.66):

$$\frac{\mathrm{d}v_y}{\mathrm{d}r} = \frac{I_Z}{m}\,\frac{F_{yF} + F_{yR} - mrv_x}{F_{yF}l_F - F_{yR}l_R} \tag{27.127}$$

As $F_{yF} = F_{yF}(\alpha_F, F_{zF})$, $F_{yR} = F_{yR}(\alpha_R, F_{zR})$, state variables α_F, α_R are more convenient. There follows from (27.68) and (27.127),

$$\begin{aligned}\frac{\mathrm{d}\alpha_R}{\mathrm{d}\alpha_F} &= \frac{(\mathrm{d}v_y/\mathrm{d}r) - l_R}{(\mathrm{d}v_y/\mathrm{d}r) + l_F} = \\ &= \frac{(I_Z/m)[F_{yF} + F_{yR} - m(v_x^2/l)(\delta_F - \alpha_F + \alpha_R)] - l_R(l_F F_{yF} - l_R F_{yR})}{(I_Z/m)[F_{yF} + F_{yR} - m(v_x^2/l)(\delta_F - \alpha_F + \alpha_R)] + l_F(l_F F_{yF} - l_R F_{yR})}\end{aligned} \tag{27.128}$$

This relation and respective axle characteristics can now be used to plot trajectories in the $\alpha_F - \alpha_R$ phase-plane by applying isoclines. Also, time integration of the equations of motion for various initial conditions yields the phase-plane diagram, which is displayed in Figure 27.31 corresponding to Figure 27.30 for $v_x \approx v = 20$ m/s and constant $\delta_F^* = 0.025$ rad. The course of the trajectories is indicated by arrows.

Setting numerator and denominator in (27.128) to zero results in governing equations for isoclines with horizontal and vertical tangents to the trajectories; setting the fraction to one yields the isocline with 45° tangents and $\dot{r}$ will vanish according to (27.64). The three isoclines are shown as dotted lines in Figure 27.31, and the intersections of them give singular points, where $\mathrm{d}\alpha_F = \mathrm{d}\alpha_R = 0$ and the motion is in a steady state, $\dot{\alpha}_F = \dot{\alpha}_R = \dot{v}_y = \dot{r} = 0$. The three equilibrium points I, II, and III in the figure correspond to the same as in Figure 27.30. Point I is a stable focus and represents a negative complex pair of eigenvalues of the characteristic equation of the system linearized at that respective point; points II and III are unstable saddle points corresponding to a real pair of eigenvalues, one of which is positive. Further, the different types of attraction of trajectories starting from inside or outside the domain bounded by indicated

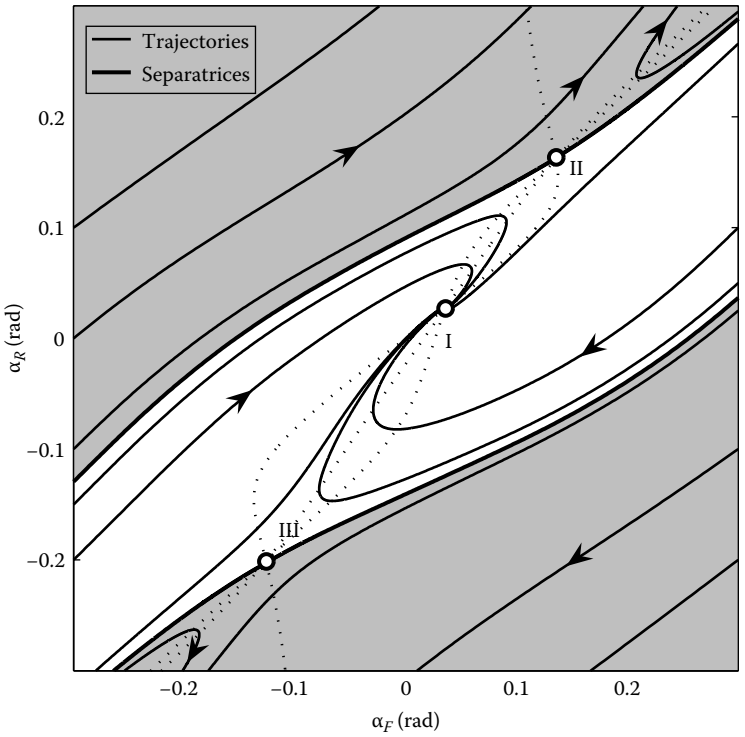


FIGURE 27.31 Phase-plane diagram corresponding to the handling diagram of Figure 27.30 for $v = 20$ m/s and $\delta_F^* = 0.025$ rad resulting in the same singular points I, II, and III.

separatrices can be observed. In this example, the trajectories starting from outside this region (in the shaded area) tend toward very large side slip angles, corresponding to an unstable motion, and trajectories from inside the domain run to the steady-state motion. However, this may not be generalized, in particular, when the front axle/tire characteristics drops below the rear axle/tire characteristics at high side slip angles. Anyway, only the domain within the separatrices is normally of interest from a technical point of view. For limited disturbances, a stability margin may be derived and the effects of different axle/tire characteristics and other parameters studied.

27.4 SUBSYSTEMS AND VEHICLE–TRAILER COMBINATION

With simple models, not only a basic understanding of the dynamics of the full vehicle can be obtained but they may also be used to explain mechanisms or phenomena originating in subsystems of the vehicle. Two examples thereof are presented. The chapter is concluded by a brief stability analysis of an articulated vehicle, a vehicle–trailer combination.

27.4.1 Braking of a Wheel until Lock-Up

Locked tires of a vehicle may not only elongate its stopping distance but stability of the motion of the vehicle can be lost as well. Only very limited lateral tire forces can be transferred ($F_y \approx \mu_s F_z \sin \alpha$), and thereby maneuverability of the vehicle is reduced considerably. Consequently, a full lock-up of wheels needs to be avoided, which is the objective of an ABS. Basically, this system recognizes a potentially locking wheel by the preceding rapid increase in the angular acceleration of the wheel, which will be revealed here, following [7].

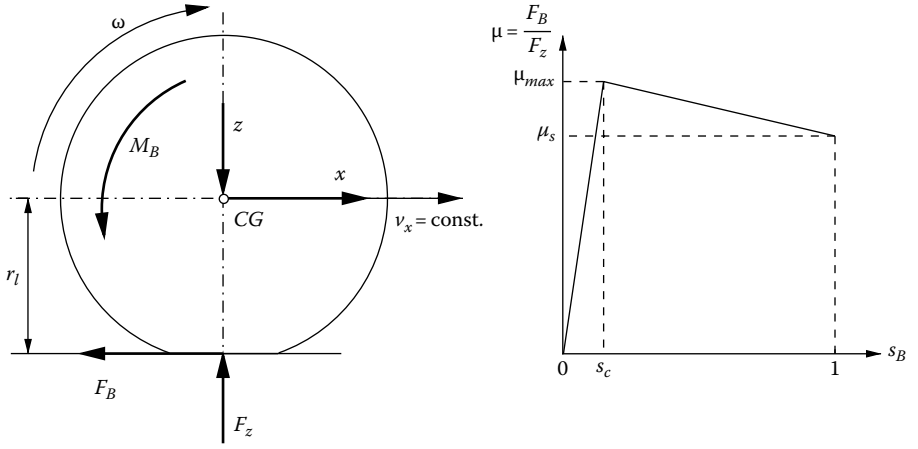


FIGURE 27.32 Forces and moments acting on the braked wheel; bilinear characteristics of braking force coefficient μ.

Euler's equation for the wheel according to Figure 27.32 is

$$I\dot{\omega} = -M_B + F_B r_l \tag{27.129}$$

with the momentum of inertia I of the wheel for the spin axis. The applied braking moments $M_B = c_B t$ is considered to increase linearly with time t until wheel lock-up. This period is presumed to be short and the decrease in velocity can be neglected; thus, v_x = constant. Also, the tire load F_z is assumed constant.

To allow for an analytical solution of the equation of motion (27.129), the braking force coefficient $\mu = F_B/F_z$ is given bilinear characteristics (see Figure 27.32), with respect to the braking slip

$$s_B = \frac{v_x - r\omega}{v_x} = 1 - \frac{\omega}{\omega_0} \tag{27.130}$$

where r is the effective rolling radius of the tire at zero slip, $v_x = r\omega_0$. Two cases are defined, when substituting bilinear $F_B = \mu(s_B)F_z$ in (27.129):

$0 \le s_B \le s_c$, $0 \le t \le t_c$:

The equation of motion for the wheel now reads

$$\dot{\omega} + \frac{\mu_{max} F_z r_l}{s_c I \omega_0}\omega = \frac{\mu_{max} F_z r_l}{s_c I} - \frac{c_B}{I} t \tag{27.131}$$

and the analytical solution becomes

$$\frac{\omega(t)}{\omega_0} = 1 - \frac{c_B s_c}{\mu_{max} F_z r_l}\left[t - \frac{s_c I \omega_0}{\mu_{max} F_z r_l}\left(1 - e^{\frac{-\mu_{max} F_z r_l}{s_c I \omega_0} t}\right)\right] \tag{27.132}$$

It can be seen from $\omega(t)$ in the numerical example in Figure 27.33 that the exponential function can be neglected after a transition period; then the period t_c to reach μ_{max}, the corresponding braking slip $s_c = 1 - \omega_c/\omega_0$, respectively, can be found as follows:

$$t_c \approx \frac{s_c I \omega_0}{\mu_{max} F_z r_l} + \frac{\mu_{max} F_z r_l}{c_B} \tag{27.133}$$

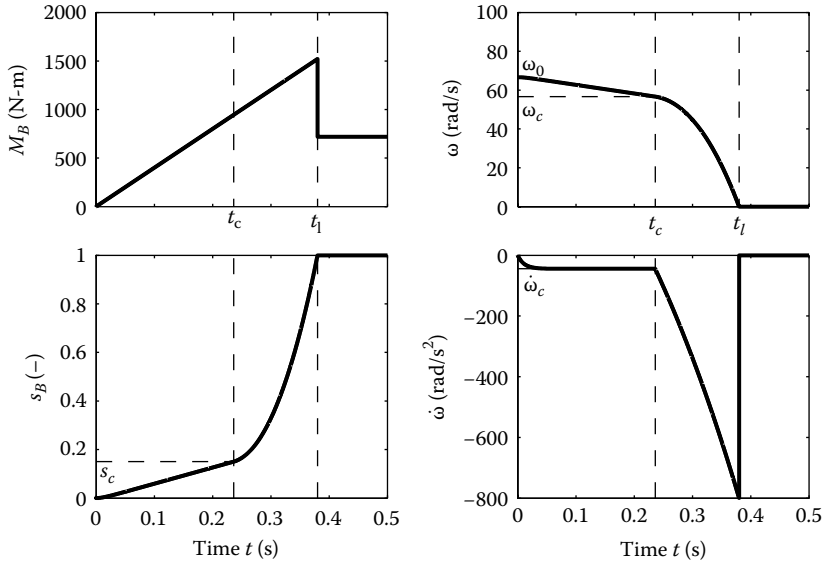


FIGURE 27.33 Braking a wheel until lock-up; model parameters $I = 1$ kg m^2, $F_z = 3000$ N, $c_B = 4000$ N-m/s, $r_1 \approx r = 0.3$ m, $v = 20$ m/s, $s_c = 0.15$, $\mu_{max} = 1$, $\mu_s = 0.8$.

And after derivation of (27.132) w.r.t. time, the corresponding angular acceleration $\dot{\omega}_c$ at t_c becomes

$$\dot{\omega}_c \approx -\frac{s_c c_B}{\mu_{max} F_z r_l}\omega_0 \tag{27.134}$$

The decrease from ω_0 toward ω_c is almost linear and therefore $\dot{\omega} \approx \dot{\omega}_c \approx$ constant before t_c.

The period between t_c and t_l is determined by the second case:

$s_c \le s_B \le 1$, $t_c \le t \le t_l$:

The analytical solution of the corresponding Euler's equation

$$\dot{\omega} - \frac{(\mu_{max} - \mu_s) F_z r_l}{(1 - s_c) I \omega_0}\omega = -\left[\frac{c_B s_c \omega_0}{\mu_{max} F_z r_l} + \frac{(\mu_{max} - \mu_s) F_z r_l}{I} + \frac{c_B}{I}(t - t_c)\right] \tag{27.135}$$

is

$$\frac{\omega(t)}{\omega_0} = \frac{(1 - s_c) c_B}{(\mu_{max} - \mu_s) F_z r_l}\left[\frac{(\mu_{max} - \mu_s) F_z r_l}{c_B} + (t - t_c) + \frac{I(\mu_{max} - s_c \mu_s)\omega_0}{\mu_{max}(\mu_{max} - \mu_s) F_z r_l}\left(1 - e^{\frac{(\mu_{max} - \mu_s) F_z r_l}{(1 - s_c) I \omega_0}(t - t_c)}\right)\right] \tag{27.136}$$

In this case and presuming $(\mu_{max} - \mu_s) > 0$, the exponential function accounts for a very quick drop in the angular velocity, with very large angular acceleration in particular, until they are both zero at t_l; when the wheel is fully locked-up, $s_B = 1$. At the locked wheel, the braking moments is not set by the driver anymore, but by the sliding tire, $M_B = r_l F_B = r_l \mu_s F_z$.

As a consequence, the ABS needs to detect angular accelerations larger than a specific limit beyond $\dot{\omega}_c$, when the wheel tends to lock-up quickly and braking pressure needs to be released.

If $s_c < s_B < 1$, the corresponding eigenvalue of (27.135) is positive, revealing a possible unstable motion. In particular, when measuring steady-state tire characteristics on a tire test bench, the unstable range $\mu_{max} < \mu < \mu_s$ induces a fast transition from μ_{max} to μ_s, as the tire cannot be operated in stable (quasi) steady-state conditions within this area!

27.4.2 Wheel Shimmy

Wheel shimmy is a self-excited, oscillatory, combined lateral-yaw motion of the wheel about a (nearly) vertical steering axis (negligible caster angle) caused by the interaction between the dynamic behavior of the tire and the suspension of the wheel. This kind of (unstable) motion may occur, for example, at the front wheels of an automobile [8], at the landing gears of an aircraft [26], or as a related phenomenon in a motorcycle (called wobble mode, Chapter 35). The trailing wheel model in Figure 27.34 is the simplest model to illustrate wheel shimmy, following [8] and [26] and for continued analysis of the phenomenon.

The moments of inertia for the vertical axis at point A is denoted by $I_A = I_C + me^2$ (wheel mass m, wheel moments of inertia I_C), caster length with e, and linear stiffness and damping constant with c_ψ and d_ψ, respectively. The hinge point A moves along the x-axis with constant velocity v and the yaw angle ψ is a DOF of the system:

$$I_A\ddot{\psi} = -d_\psi\dot{\psi} - c_\psi\psi - F_y e - M_z \tag{27.137}$$

The straight tangent tire model is applied to account for the transient tire behavior. Thereby, the approximated shape of a single deflected string, representing the tire, is a straight line tangent to the actual contact line at the leading edge. The corresponding deformation angle α_t is another DOF. Considering small deviations from a straight-line motion, the transient tire model follows to

$$\sigma\dot{\alpha}_t + v\alpha_t = v\psi - a\dot{\psi} + \dot{y}_C \tag{27.138}$$

$$F_y = C_F\alpha_t \tag{27.139}$$

$$M_z = C_M\alpha_t \tag{27.140}$$

with $\dot{y}_C = e\dot{\psi}$, relaxation length σ, half of the tire contact length a, cornering stiffness C_F, self-aligning stiffness C_M, and consequently the pneumatic trail $t = M_z/F_y = C_M/C_F$.

The characteristic equation of the linear third-order system (27.137) through (27.140) is as follows:

$$a_0 s^3 + a_1 s^2 + a_2 s + a_3 = 0 \tag{27.141}$$

$$\text{with } a_0 = \sigma I_A \tag{27.142}$$

$$a_1 = vI_A + \sigma d_\psi \tag{27.143}$$

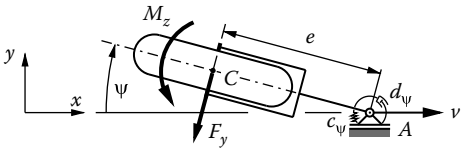


FIGURE 27.34 Trailing wheel model.

$$a_2 = \sigma c_\psi + C_F(e-a)(e+t) + v d_\psi \tag{27.144}$$

$$a_3 = v c_\psi + C_F v(e+t) \tag{27.145}$$

According to the Hurwitz (and Stodola) criterion, $a_i > 0$ (i = 0–3) and $H_2 = a_1 a_2 - a_0 a_3 > 0$, and assuming only positive parameters but the mechanical trail e, the conditions for a stable motion without damper, $d_\psi = 0$, finally result in

$$a_3 > 0: \quad e > -\left(t + \frac{c_\psi}{C_F}\right) = e_1 \tag{27.146}$$

$$H_2 > 0: \quad e < -t = e_2 \quad \text{or} \quad e > \sigma + a = e_3 \tag{27.147}$$

Thus, the system is oscillatory unstable (H_2 to become negative first, [27]) for $-t < e < \sigma + a$, independent from v, I_A, c_ψ, C_F (but changing the eigenfrequency)! If $c_\psi = 0$, the stable region for negative mechanical trail disappears. If $e < -t - c_\psi/C_F$, meaning that the steering axis lies behind the contact center, the system becomes monotonically unstable (a_3 to become negative first, [27]) and the wheel turns back around the swivel axis to a trailing motion. These analytical findings are illustrated for the numerical example in Figure 27.35, top left plot. The top right plot shows the root locus for $d_\psi = 0$ and increasing mechanical trail from $e = -0.6$ m, indicated by "*", to $e = 0.6$ m, marked with "o". Although the stability margins for e (e_1, e_2, e_3) are independent from velocity v, when no yaw damping ($d_\psi = 0$) is considered, natural frequency and degree of stability very well depend thereon, demonstrated by the two root loci for v = 15 m/s and v = 30 m/s. Note, all root loci for different velocities cross the imaginary axis at the same points.

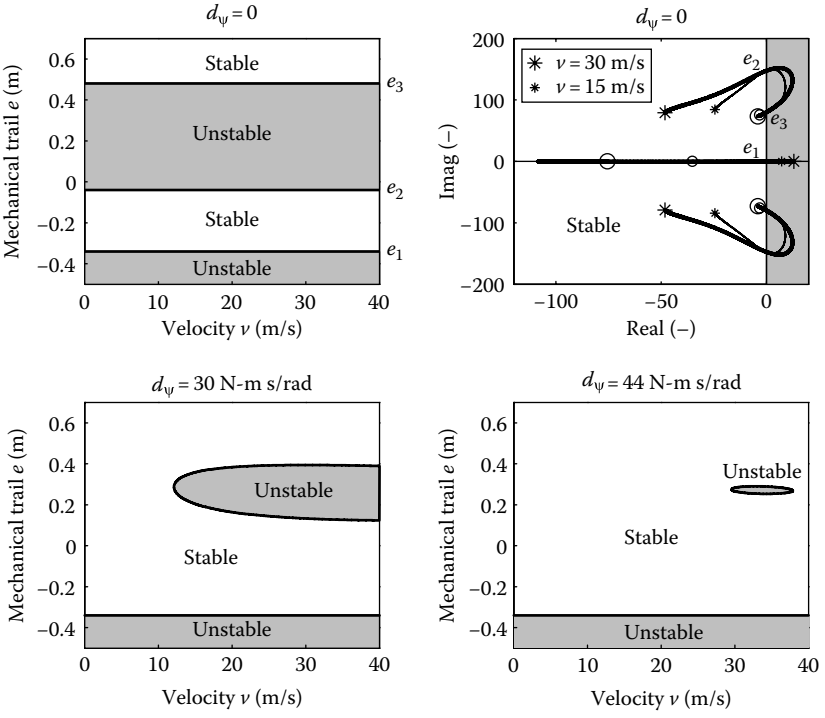


FIGURE 27.35 Stability map of the trailing wheel system in Figure 27.34 for $d_\psi = 0$, d_ψ = 30 N-m s/rad, d_ψ = 44 N-m s/rad; root locus for −0.6 m (*) < e < 0.6 m (o); model parameters I_C = 0.65 kg m², m = 25 kg, C_F = 50,000 N/rad, c_ψ = 15,000 N-m/rad, σ = 0.36 m, a = 0.12 m, t = 0.04 m.

By introducing yaw damping (bottom plot in Figure 27.35), the unstable area between $-t < e < \sigma + a$ for $d_\psi = 0$ can be reduced and disappears completely, when damping is sufficiently high. Nevertheless, the unstable area for large negative mechanical trails $e < -t - c_\psi / C_F$ remains. For a given trail, the stability of the motion may now depend on velocity!

27.4.3 Vehicle–Trailer Combination

A final example for a basic stability analysis is the motion of a vehicle–trailer combination. The linear two-wheel vehicle model in Figure 27.18 is coupled with a model of the trailer based on the same level of complexity and modeling assumptions (Figure 27.36).

The 2 DOFs of the towing vehicle, lateral velocity $v_y = v\beta$ and yaw rate $\dot{\psi} = r$ of the vehicle, are completed by the articulation angle $\varphi = \psi - \psi_T$ between the vehicle and trailer. The velocity $v \approx v_x$ of the towing vehicle is considered a constant parameter. Then, the linear system equations in matrix notation read

$$\mathbf{E}\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\delta_F \tag{27.148}$$

with state vector $\underline{x} = [v_y, r, \varphi, \dot{\varphi}]^T$, front-wheel steering angle δ_F, and matrices (with trailer mass, m_T, moments of inertia, I_T, and cornering stiffness, C_T)

$$\mathbf{A} = \begin{bmatrix} -(C_F + C_R + C_T)\dfrac{1}{v} & -(m + m_T)v - C_F\dfrac{l_F}{v} + C_R\dfrac{l_R}{v} + C_T\dfrac{h+b}{v} & -C_T & -C_T\dfrac{b}{v} \\ -C_F\dfrac{l_F}{v} + C_R\dfrac{l_R}{v} + C_T\dfrac{h}{v} & m_T h v - C_F\dfrac{l_F^2}{v} - C_R\dfrac{l_R^2}{v} - C_T\dfrac{h(h+b)}{v} & C_T h & C_T\dfrac{bh}{v} \\ 0 & 0 & 0 & 1 \\ -C_T\dfrac{b}{v} & -m_T a v - C_T\dfrac{b(b+h)}{v} & -C_T b & -C_T\dfrac{b^2}{v} \end{bmatrix} \tag{27.149}$$

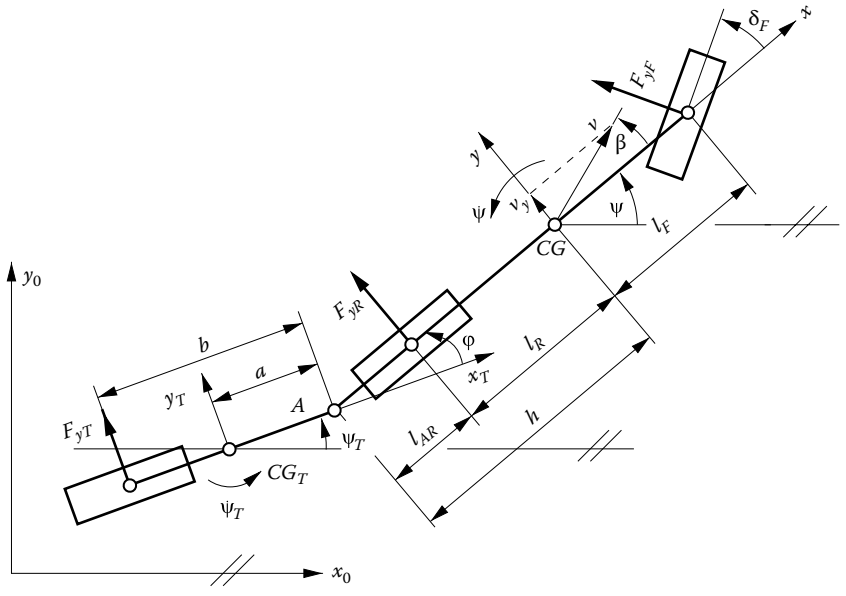


FIGURE 27.36 Vehicle–trailer model.

$$\mathbf{B} = \begin{bmatrix} C_F \\ C_F l_F \\ 0 \\ 0 \end{bmatrix} \tag{27.150}$$

$$\mathbf{E} = \begin{bmatrix} (m+m_T) & -m_T(h+a) & 0 & m_T a \\ -m_T h & I_Z + m_T h(h+a) & 0 & -m_T h a \\ 0 & 0 & 1 & 0 \\ m_T a & -(I_T + m_T a^2) - m_T a h & 0 & I_T + m_T a^2 \end{bmatrix} \tag{27.151}$$

and may be used, for instance, to design a control system or to study handling characteristics or stability issues. If one is interested in the stability of the steady-state motion, the eigenvalues, resulting numerically from $|\mathbf{A} - \lambda\mathbf{E}| = 0$, may be consulted. If one requires to evaluate system parameters with respect to stability by preparing a stability map, the coefficients a_i of the characteristic equation, $a_0\lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0$, may (symbolically) be derived. Then, asymptotic stability is guaranteed for all $a_i > 0$ ($i = 0$–4) and $H_3 = a_1a_2a_3 - a_1^2a_4 - a_0a_3^2 > 0$ (Liénard-Chipart and Stodola criterion).

In Figure 27.37, stability maps for a small vehicle trailer model, data are listed in the caption of the figure, towed by vehicle A with parameters from Table 27.1, are shown. The additional load on the rear facilitates a towing vehicle with oversteer characteristics; therefore, this type of vehicle has been selected.

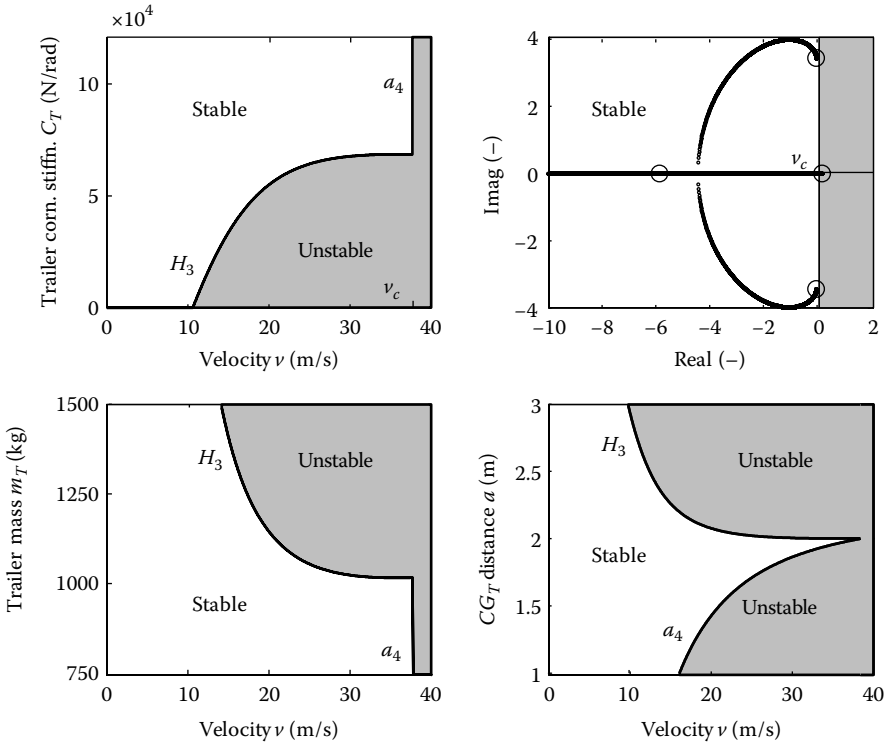


FIGURE 27.37 Stability maps of vehicle–trailer model in Figure 27.36; towing vehicle: vehicle A according to Table 27.1; trailer: $m_T = 1000$ kg, $I_T = 1{,}440$ kg m^2, $C_T = 70{,}000$ N/rad, $a = 2$ m, $b = 2$ m, $h = 2$ m; root locus for $v \to 40$ m/s (marked by "0").

From the top left plot in Figure 27.37, it can be concluded that decreasing the cornering stiffness of the trailer promotes oscillatory instability. Note, if the parameters are varied beginning within the region of asymptotic stability, the stability boundary is reached, when $a_4 H_3 = 0$, [27]. Here, $a_4 = 0$ characterizes the monotone marginal stability, $H_3 = 0$ the oscillatory marginal stability, and the respective stability boundaries are indicated by "a_4" and "H_3" in the plots. Thus, for low cornering stiffness of the trailer, a "snaking" oscillatory motion may be initiated after exceeding the stable region, "H_3." Large (initiated) amplitudes will increase the side slip angle, and as a consequence of the degressively, actually nonlinear lateral tire characteristic, the "average" cornering stiffness is decreased (further), encouraging snaking. As a result, the vehicle–trailer combination may even fallover [8]. For high cornering stiffness, divergent instability may occur, when velocity is increased just beyond the critical speed v_c, indicated by the vertical line, "a_4"; $v_c = 37.8$ m/s does not depend on the cornering stiffness C_T. As an example, the locus of eigenvalues is plotted for increasing velocity with fixed $C_T = 70{,}000$ N/rad (see top right plot in the figure). At $v = v_c$, monotone marginal stability is reached, followed by divergent instability until $v = 40$ m/s, marked by "0".

The left plot on the bottom of Figure 27.37 depicts an increasing stable area with decreasing total mass m_T of the trailer. It should be mentioned that the graph "a_4" does not depend on m_T in this specific case, when $a = b = h$!

Moving the load of the trailer toward the front end of the trailer (smaller distance a) supports stability, although it seems there is a kind of optimal position of the trailer's center of gravity with respect to the hinge, respectively the axle of the wheels.

Details are to be found in [8], and the stability analysis of a nonlinear vehicle–semitrailer combination is covered in [28] and [29].

REFERENCES

1. Abe M.: *Vehicle Handling Dynamics*: *Theory and Application*. Butterworth/Heinemann, Oxford, U.K., 2009.
2. Ammon D.: *Modellbildung und Systementwicklung in der Fahrzeugdyamik*. Teubner Verlag, Stuttgart, Germany, 1997.
3. Bastons C., Howard G.: *Car Suspension and Handling*. Pentech Press, London, U.K., 1993.
4. Dixon J.C.: *Tires, Suspension and Handling*. SAE Inc., Warrendale, PA, 1996.
5. Ellis J.R.: *Vehicle Handling Dynamics*. MEP, London, U.K., 1994.
6. Gillespie T.D.: *Fundamentals of Vehicle Dynamics*. SAE Inc., Warrendale, PA, 1992.
7. Mitschke M., Wallentowitz H.: *Dynamik der Kraftfahrzeuge*. Springer Verlag, Berlin, Germany, 2004.
8. Pacejka H.B.: *Tyre and Vehicle Dynamics*. Butterworth/Heinemann, Oxford, U.K., 2012.
9. Popp K., Schiehlen W.: *Ground Vehicles Dynamics*. Springer Verlag, Berlin, Germany, 2010.
10. Rill G.: *Road Vehicles Dynamics*: *Fundamentals and Modeling*. CRC Press, Taylor & Francis Group, Boca Raton, FL, 2012.
11. Willumeit H.-P.: *Modelle und Modellierungsverfahren in der Fahrzeugdynamik*. Teubner Verlag, Stuttgart, Germany, 1998.
12. *Bosch Automotive Handbook*, 6th edn. Robert Bosch GmbH Distribution, SAE, Warrendale, PA, 2004.
13. Lugner P.: Plöchl M.: Dynamic traction limitations of a passenger car with 4-wheel drive. *International Journal of Vehicle Design*, 25(3), 189–197, 2001.
14. Wong J.Y.: *Theory of Ground Vehicles*, 3rd edn. John Wiley & Sons, New York, 2001.
15. Hucho W.H. (ed.): *Aerodynamik des Automobils*. Vogel Verlag, Würzburg, Germany, 1999.
16. Milliken W.F., Milliken D.L.: *Race Car Vehicle Dynamics*. SAE International, Warrendale, PA, R 146, 1995.
17. SAE Handbook Supplement: *Vehicle Dynamics Terminology*. Society of Automotive Engineers Inc., Warrendale, PA, last revision 1976.
18. Lugner P., Plöchl M., Desoyer K.: *Skriptum zur Vorlesung "Fahrzeugdynamik"*. Institut für Mechanik und Mechatronik, Technische Universität Wien, Vienna, Austria, 2004.
19. Genta G.: *Motor Vehicle Dynamics*. World Scientific, Singapore, 1997.
20. Kortüm W., Lugner P.: *Systemdynamik und Regelung von Fahrzeugen*. Springer Verlag, Berlin, Germany, 1994.

21. Lugner P.: Horizontal motion of automobiles: Theoretical and practical investigations. CISM Courses and Lectures No 274, W.O. Schiehlen (ed.), Springer Verlag, New York, 1982.
22. Zomotor A.: *Fahrwerktechnik: Fahrverhalten*. Reimpell J. (ed.). Vogel Buchverlag, Würzburg, Germany, 1987.
23. Furukawa Y. et al: A review of four-wheel steering studies from the viewpoint of vehicle dynamics and control. *Vehicle System Dynamics*, 18: 151–186, 1989.
24. Pacejka H.B. (ed.): *1st International Colloquium on Tyre Models for Vehicle Dynamics Analysis*. Delft, the Netherlands. *Vehicle System Dynamics*, 21(S1), 1993.
25. Winkler Ch.B.: 1st simplified analysis of the steady-state turning of complex vehicles. *Vehicle System Dynamics*, 29: 141–180, 1998.
26. Besselink I.J.M.: Shimmy of aircraft main landing gears. PhD. thesis, Delft University of Technology, Delft, the Netherlands, 2000.
27. Müller P.C., Schiehlen W.: *Linear Vibrations*. Martinus Nijhoff Publishers, Dordrecht, the Netherlands, 1985, p. 110.
28. Troger H., Zeman K.: A nonlinear analysis of the generic types of loss of stability of the steady state motion of a tractor-semitrailer. *Vehicle System Dynamics*, 13: 161–172, 1984.
29. Troger H., Steindl A.: *Nonlinear Stability and Bifurcation Theory: An Introduction for Engineers and Applied Scientists*. Springer-Verlag, New York, 1991.

28 Detailed Modeling, Simulation, and Analysis of Vehicle Dynamics

Dieter Ammon

CONTENTS

28.1 Introduction 1024
28.1.1 Model-Based Vehicle Dynamics and Chassis Development 1024
28.1.2 Vehicle Dynamics and Chassis Simulation 1025
28.1.3 Coordinate Systems for the Description of Vehicle Movement 1025
28.1.4 Wheel Motions 1027
28.2 Road Modeling 1028
28.2.1 Deterministic Road Irregularities 1028
28.2.2 Stochastic Road Irregularities 1029
28.2.3 Road Characteristics in Transverse Direction 1030
28.2.4 Road Height Profile Modeling 1031
28.2.5 Synthesis of Two-Track Excitations 1032
28.3 Tire Models 1032
28.3.1 Application Scope and Related Demands/Requirements 1033
28.3.2 Road, Tire, and Vehicle Interfaces 1033
28.3.3 Tire Model Parameterization 1033
28.3.4 Tire Dynamics and Low-Speed Properties 1035
28.3.5 Friction Behavior and Modeling 1036
28.3.6 Tire Comfort Behavior and Modeling 1037
28.3.7 Challenges in Future Tire Modeling 1038
28.4 Axles: Chassis Components 1039
28.4.1 Chassis Mounts and Bearings 1039
28.4.2 Elastokinematic Adjustment 1042
28.5 Suspension Systems 1043
28.5.1 Steel-Spring Suspension and Suspension Characteristics 1044
28.5.2 Air Suspension and Air Suspension Modeling 1044
28.5.3 Dampers and Damper Modeling 1045
28.5.4 Variable Damper Systems 1047
28.5.5 Push and Pull Stops 1047
28.5.6 Active Suspension Systems 1047
28.6 Steering Systems 1049
28.6.1 Power Steering Systems 1049
28.6.2 Transfer Characteristics 1051
28.6.3 Steering Models for Vehicle Behavior Analysis 1051
28.6.4 Detailed Steering System Models for Axle Dynamics Analysis 1052
28.6.5 Active Steering Systems 1053
References 1055

28.1 INTRODUCTION

The vehicle dynamics design and adjustment of present-day road vehicles are intended to ensure the following:

- The vehicle can be controlled with precision and is robust with regard to disturbances.
- The driver can master practically all driving situations.
- Vehicle responses are predictable for the driver.
- Comfort, traction, stability, and, if applicable, off-road characteristics appropriate to the vehicle type can be realized.

In order to be able to comply with these ever-more complex requirements, the use of efficient and sometimes complex analytical tools is indispensable [1–3]. This applies, in particular, to the vehicle and system design, but increasingly also to the support and optimization of test series and to the appropriate adjustment for the customer [4,6].

28.1.1 Model-Based Vehicle Dynamics and Chassis Development

Within the context of vehicle and chassis development, analysis by means of simulation technology forms the starting point and backbone of the development process, as illustrated in Figure 28.1.

The analysis using modeling and simulation allows the functional properties and experiences from the relevant predecessor or reference vehicle to be compiled in a development model and to be configured specific to the target vehicle. New or heavily modified components are analyzed by means of rig tests, prepared in line with the model technology and integrated into the overall system analysis.

If technical innovations impact upon the interaction between the driver and vehicle, these aspects can be dealt with within the framework of driving simulator evaluations and preoptimized long before elaborate overall vehicle hardware is available [8].

Based on this preliminary work, a large proportion of the necessary design and adjustment issues can be resolved systematically by means of simulation technology; this applies, in particular, to the functioning and robustness of the design and the remaining adjustment scope. The concluding prototype phase is primarily intended not only for the validation of the results under real driving

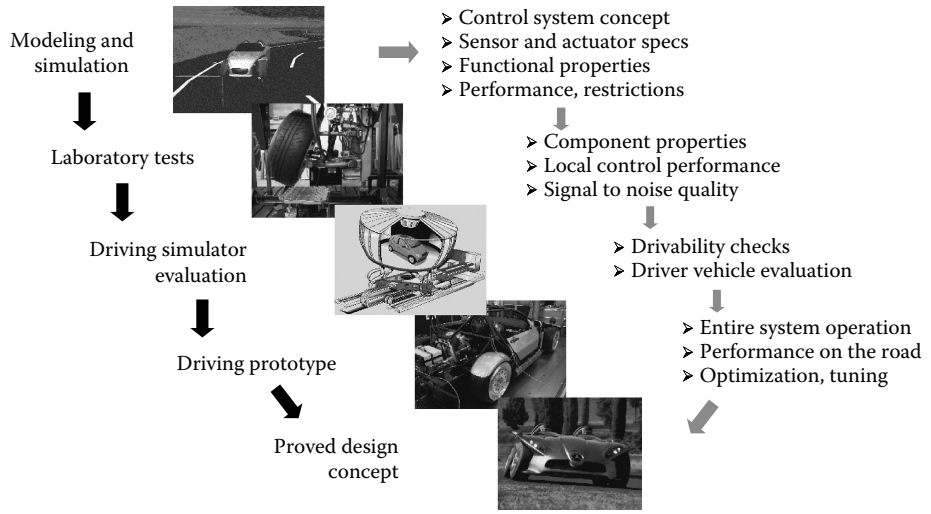


FIGURE 28.1 Process concept of a model-based vehicle dynamics and chassis development.

conditions but also for checking and fine tuning, as the technical evaluation criteria available today still only cover part of the human perception. Vehicle evaluation and sensitive adjustments by experienced test drivers are therefore indispensable parts of the development process.

All in all, at the end of the development process chain, we have a vehicle or system concept that is mature with regard to function, robustness, and adjustment options and which therefore forms the basis for low-risk and timely industrial realization.

28.1.2 Vehicle Dynamics and Chassis Simulation

The vehicle dynamics as a whole involve the complex interaction between driver and vehicle, including the integrated vehicle dynamics, ride comfort, and driving safety systems, as well as the disturbance and control behavior of the vehicle in relation to the vehicle environment, as illustrated in Figure 28.2.

Given the complexity of the vehicle dynamics analysis and the heterogeneity of the tasks in hand, three main model categories of differing complexity have evolved for the description and explanation of vehicle dynamics phenomena and the solution of adjustment problems (refer to Figure 28.3):

- Vehicle behavior models
- Handling models
- Detailed ride analysis models (including axle dynamics properties)

Vehicle behavior models such as single-track (bicycle models) models or quarter-vehicle models each approximate certain aspects of vehicle behavior, for example, the low-frequency lateral dynamics or the suspension characteristics and the disturbance transfer of road irregularities.

In contrast, handling and chassis dynamics models for the complete vehicle are prepared and are suitable for the analysis of any driving situation. Both model classes are designed for driving on any type of road; they are designed to provide plausible results for any type of (mechanically compatible) driver input and to describe the vehicle behavior in the relevant validation range with defined quality/precision.

28.1.3 Coordinate Systems for the Description of Vehicle Movement

The spatial movement of the vehicle body or structure is described based on a vehicle-fixed coordinate system *B* (*B*: Body), that is, which moves along with the vehicle (refer to Figure 28.4).

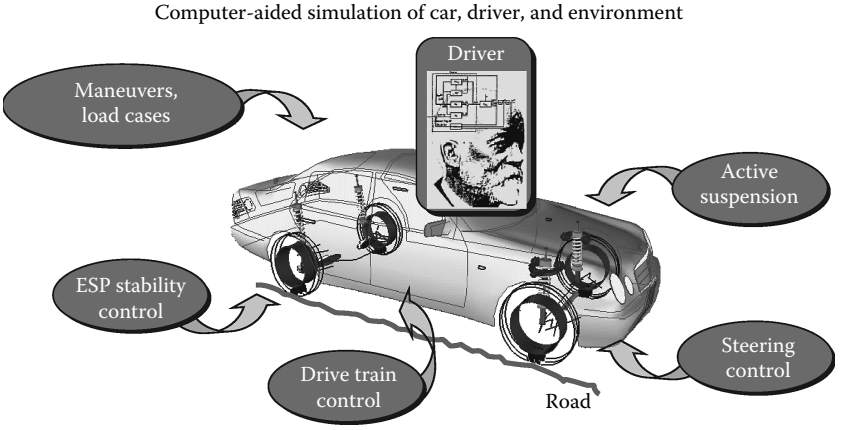


FIGURE 28.2 Interaction between driver, vehicle, and environment within the context of evaluation of vehicle dynamics.

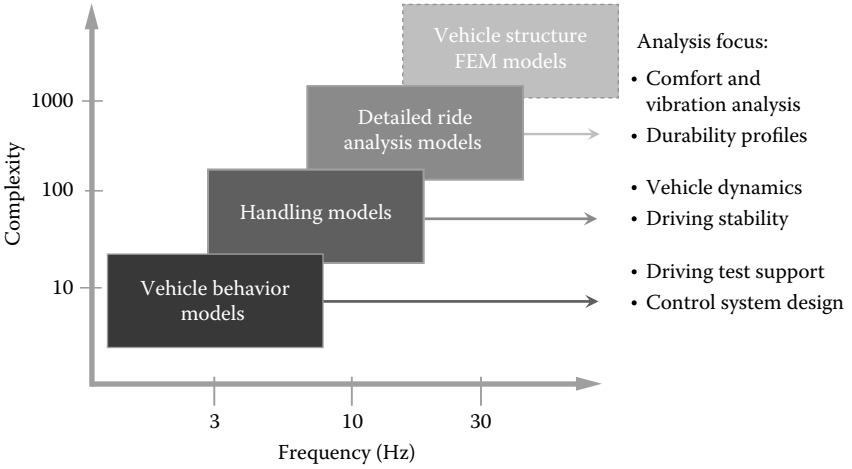


FIGURE 28.3 Model classes for vehicle dynamics analysis, arranged according to complexity and main primary analysis frequency range.

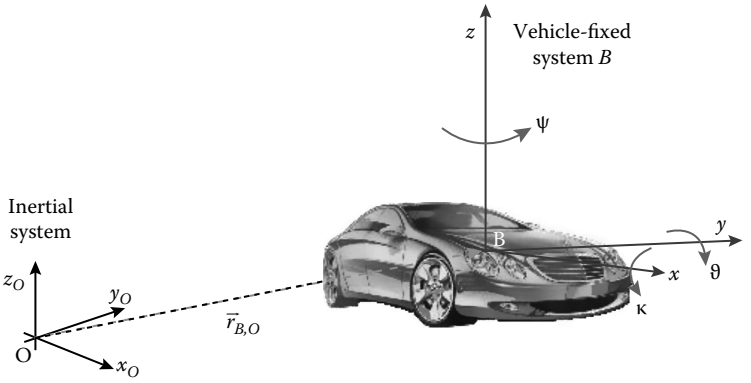


FIGURE 28.4 Description of vehicle motions in space: vehicle-fixed system *B*, aligned to the longitudinal, lateral, and vertical vehicle axis.

The vehicle center of gravity—at a defined reference load—lends itself as reference point. Although this does not simplify the dynamics equations significantly, it facilitates the interpretation of the analytical results. The vehicle body motion is characterized by the inertial position and speed values of reference point *B*

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \vec{r}_{B,O}; \quad \vec{v} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \vec{v}_{B,O} \lim_{x \to \infty} \tag{28.1}$$

and the angle and angular velocity values of the consecutive body rotations around the vertical, lateral, and longitudinal axis:

$$\vec{\varphi} = \begin{pmatrix} \kappa \\ \vartheta \\ \psi \end{pmatrix}; \quad \dot{\vec{\varphi}} = \begin{pmatrix} \dot{\kappa} \\ \dot{\vartheta} \\ \dot{\psi} \end{pmatrix}; \quad \vec{\omega} = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\sin\vartheta \\ 0 & \cos\kappa & \sin\kappa\cos\vartheta \\ 0 & -\sin\kappa & \cos\kappa\cos\vartheta \end{pmatrix} \dot{\vec{\varphi}} \tag{28.2}$$

Using reference point values $\vec{r}_B$, $\vec{v}_B$, and a rotation matrix $\bar{T}$, all relevant quantities of body system B can be transformed into the inertial system O and vice versa, for example,

$$\vec{r}_{P,O} = \vec{r}_{B,O} + \bar{T}\Delta\vec{r}_{P,B}; \quad \vec{v}_{P,O} = \vec{v}_{B,O} + \bar{T}\Delta\vec{v}_{P,B} + \vec{\omega}\times\bar{T}\vec{r}_{P,B} \tag{28.3}$$

The rotation matrix $\bar{T}$ results from consecutive rotations of the basic system around the coordinate axes:

$$\bar{T} = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\vartheta & 0 & \sin\vartheta \\ 0 & 1 & 0 \\ -\sin\vartheta & 0 & \cos\vartheta \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\kappa & -\sin\kappa \\ 0 & \sin\kappa & \cos\kappa \end{pmatrix} \tag{28.4}$$

28.1.4 Wheel Motions

The principal motions of the wheels relative to the vehicle body concern the degrees of freedom jounce z_e, wheel steering δ_R, and the wheel rotation φ_R, $\dot{\varphi}_R = \omega_R$, as illustrated in Figure 28.5.

These types of wheel movements are possible due to axle kinematics and enable larger displacements or angle ranges. It should be noted, however, that the basic chassis design generally provides for deviations to the main vehicle axes. The wheel axis around toe-in $\Delta\delta_R$ and camber γ_R, for example, is twisted against the transverse vehicle axis. During jounce or rebound z_e, the longitudinal and lateral position x_{RC}, y_{RC} of the wheel contact patch changes. The former results from the chassis-specific acceleration/braking compensation, while the latter corresponds with the so-called roll center height. (The roll center is generally defined as the center of rotation between the axle and vehicle under free rolling motions.)

With known wheel kinematics of a single-wheel suspension, all wheel carrier motions, that is, the longitudinal and lateral displacement as well as the wheel carrier camber γ_R and the wheel carrier pitch angle φ_{WC}, can be determined from the jounce motion z_e and the (projected) wheel steering angle δ_R:

$$x_{RC} = x_{RC}(z_e, \delta_R); \quad y_{RC} = y_{RC}(z_e, \delta_R) \tag{28.5}$$

$$\gamma_R = \gamma_R(z_e, \delta_R); \quad \varphi_{WC} = \varphi_{WC}(z_e, \delta_R) \tag{28.6}$$

In the case of rigid axles, the conditions are somewhat more complex. Here, four-dimensional characteristics are generally required, as the jounce and steering status of both wheels of an axle influence

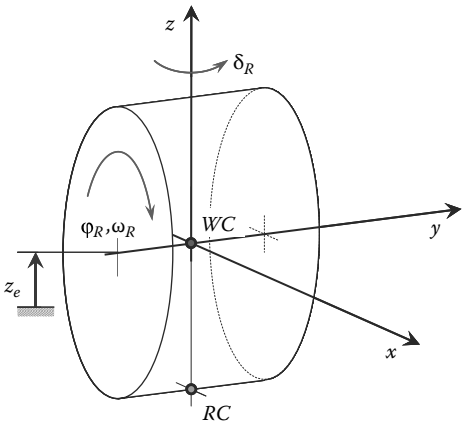


FIGURE 28.5 Primary kinematic wheel degrees of freedom, here as relative coordinates defined in relation to the vehicle system B (*RC*, wheel contact patch; *WC*, wheel center point).

the wheel guidance. Three-dimensional characteristics frequently suffice, however, as the steering movements between the left-hand and right-hand axle side are normally kinematically coupled.

Besides these major wheel motions, force and torque balances determine the actual wheel position in other directions (refer to Section 28.4).

28.2 ROAD MODELING

Driving comfort analyses as well as spring–damper adjustments and the design of semi-active and active suspensions must take road excitations into account. In practice, this is either achieved on the basis of deterministic test obstacles or specified road profiles or by measured or synthetically generated "stochastic" road profiles.

28.2.1 Deterministic Road Irregularities

The spectrum of deterministic irregularity excitations ranges from short, high-frequency irregularity excitations such as cleats and simulated slab joints (vertical steps) via simulations of potholes and speed bumps right thru to extended ramp types (along the straight or in corners) to produce stronger body excitations (refer to Figure 28.6). Typical disturbances in the transverse road profile include lateral slopes or lane groove profiles. Road-building-related road characteristics and profiles with their customary transition spirals (clothoids) between different curvatures, related transverse slopes, and curve banking as well as incline and descent profiles are also normally assigned to the deterministic road excitations. By making minor additions, the road construction descriptions also allow the definition of racetracks and high-bank curve sections. The latter are particularly important for the assurance of driving dynamics control systems, as they allow potential weak points in the estimation of the driving condition at more steeply inclined locations to be uncovered.

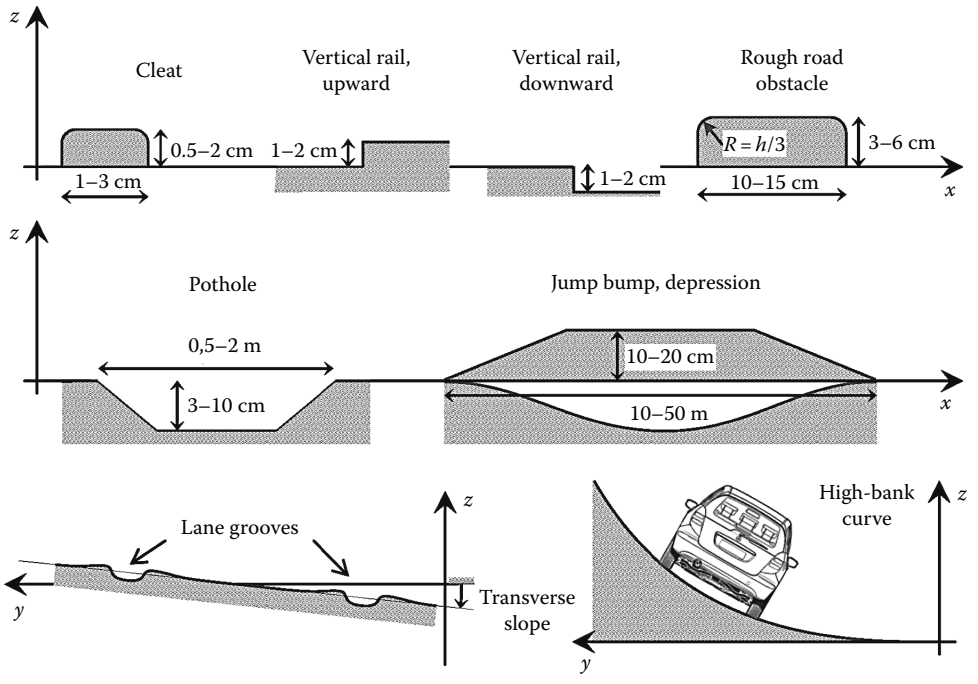


FIGURE 28.6 Selection of deterministic road irregularity excitations for ride comfort and vehicle dynamics analyses.

28.2.2 Stochastic Road Irregularities

The construction and wear-related irregularities of country roads and highways fit in quite well with the model of a random process ξ_s, which is in steady state over travel distance s and shows a Gaussian distribution. (In stochastic processes, the ordinal variable is normally used as index number—in contrast to the bracketed notation for deterministic processes.) The knowledge of the related power spectral density $S_{\xi\xi}(\Omega)$ therefore suffices to describe the road characteristics completely in longitudinal direction [16,20].

A number of analyses [12,18,20] prove that most road surfaces are approximated quite well by spectra of the type

$$S_{\xi\xi}(\Omega) = S_0 \left(\frac{\Omega_0}{\Omega} \right)^w ; \quad \Omega_0 = \frac{1}{m} \tag{28.7}$$

In double-logarithmic representation, this expression results in a straight line. The amplitude parameter S_0 therefore assumes values between 1 and 200 cm^3. For the "waviness" parameter (slope in double-logarithmic-scaled diagrams), values in the range of w = 1.8–3.8 have been determined, whereby larger waviness values have been found for very good roads and smaller amplitude parameters S_0. At a given vehicle speed v, the road spectrum can be transformed into an equivalent time-domain power spectral density (Figure 28.7):

$$S_{\xi\xi}(\omega) = \frac{S_0}{v} \left(\frac{\Omega}{\Omega_0 v} \right)^{-w} ; \quad \text{or} \quad \text{for } w = 2\text{: } S_{\xi\xi} = \frac{v S_0 \Omega_0^2}{\omega^2} \tag{28.8}$$

As can be seen for waviness w = 2, the power input of the road excitation increases proportionally to the vehicle speed; the excitation amplitudes therefore increase with $\sqrt{v}$. Above the excitation frequency, the intensity decreases quadratically. If instead of the road excitation ξ_t we observe its formal time derivative $\dot{\xi}_t$, a uniform white spectrum $S_{\dot{\xi}\dot{\xi}} = v S_0 \Omega_0^2$ results for w = 2. This is therefore also known as white speed noise. Speed-sensitive chassis parts such as shock absorbers or chassis

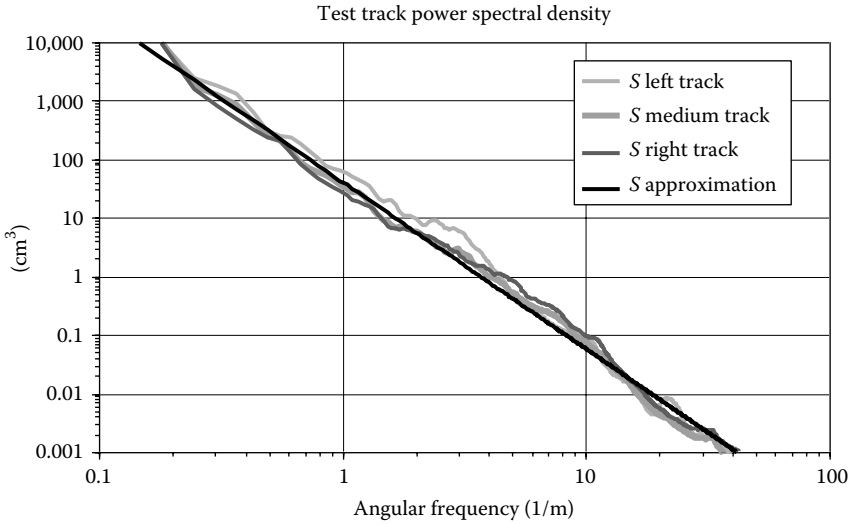


FIGURE 28.7 Power spectral density of a typical comfort evaluation testing ground, different tracks.

mounts are therefore subject to very broadband excitations. Altogether, the road excitation affects all frequency bands equally—each potential resonance in the chassis or in the overall vehicle is reliably excited.

From a theoretical point of view, model spectra of types (28.7) or (28.8) suffer from a severe deficit: The related processes ξ_s are not defined, and their variance is infinitely large. In practice, this problem hardly surfaces, however, as the disturbing, extreme low-frequency excitation content—for example, for $\Omega \leq 0.01$/m—is blanked out or—and this has the same effect—only distances of some 100 m length are used for the analysis. (The valleys and peaks along the natural road course are therefore removed from the stochastic part of the road modeling and assigned to the deterministic profile content.)

28.2.3 Road Characteristics in Transverse Direction

The road characteristics in transverse direction and the correlations in the excitations of the left and right track (or vehicle side) are described by the coherence γ of the excitations [15,18]:

$$\gamma(\rho;\Omega) = \frac{S_{lr}(\Omega)}{\sqrt{S_{ll}(\Omega)\,S_{rr}(\Omega)}}. \tag{28.9}$$

where

indices *r*,*l* are the right and left track of the vehicle

ρ is the track width

Figure 28.8 shows the typical coherence function characteristics for different track widths.

As the database for real coherence characteristics is still weak and as the theory of two-dimensional stochastic processes, on the other hand, does not leave much leeway when the longitudinal spectra are defined, it is expedient to use a synthesis of the not quite applicable isotropic model results and the plausible road data evaluations for determining coherence. This can be achieved most easily if the coherence is initially determined conforming to the isotropic model and then somewhat compressed or stretched in longitudinal road direction in order to model the longitudinal

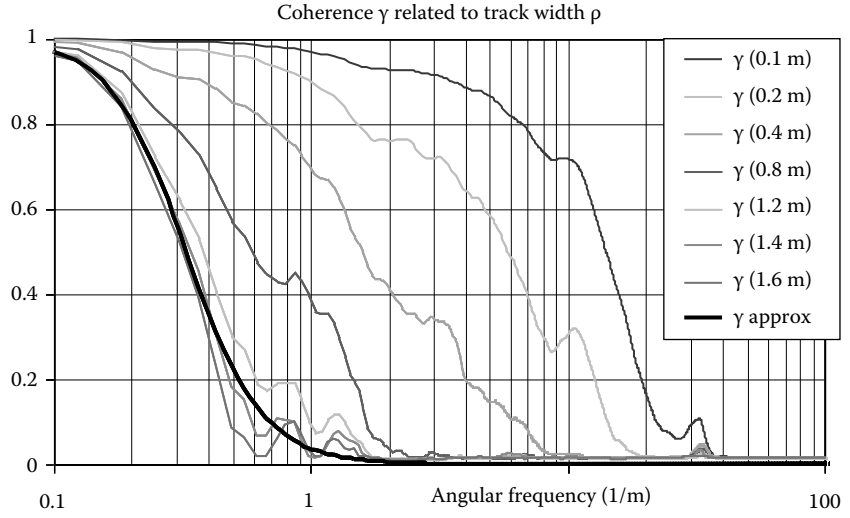


FIGURE 28.8 Coherence functions of a typical comfort evaluation testing ground for different track widths.

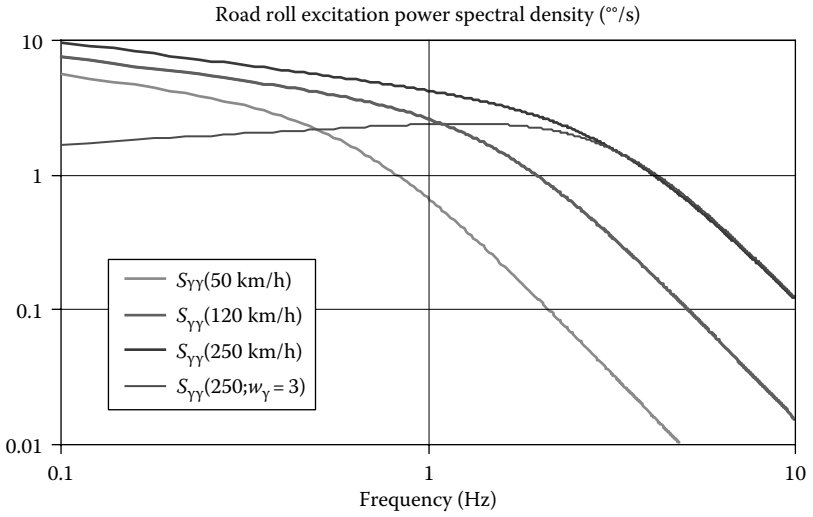


FIGURE 28.9 Roll excitation spectra for different vehicle speeds determined for waviness w = 2.84 and coherence parameter Ω_γ = 0.3/m, w_γ = 2.5.

orientation of the individual disturbances observed in practice. The results of this process are coherence approximations of the type

$$\gamma = \frac{\Omega_\gamma^{w_\gamma}}{\Omega_\gamma^{w_\gamma} + \Omega^{w_\gamma}}; \quad \Omega_\gamma \approx \frac{0{,}5 \cdots 1}{\rho}; \quad w_\gamma = w_\gamma(\rho) \approx (0{,}5 \cdots 1{,}5) w \tag{28.10}$$

The practical significance of the coherence function becomes evident, among others, when we look at the development of the roll excitations over the vehicle speed based on the aforementioned principles. The relevant power spectral density $S_{\gamma\gamma}(\omega)$ results as follows:

$$S_{\gamma\gamma}(\omega) = \frac{2}{\rho^2} \frac{S_0}{v} \left(\frac{\omega}{\Omega_0 v} \right)^{-w} \left(1 - \gamma\left(\rho; \frac{\omega}{v} \right) \right) \tag{28.11}$$

Figure 28.9 shows the power spectral densities of the roll excitation for different vehicle speeds v. It becomes clear that significant roll excitations of the vehicle body are to be expected, above all, at higher speeds. For $w_\gamma > w$, a pronounced excitation maximum occurs near the natural vehicle roll frequency.

28.2.4 Road Height Profile Modeling

Two different approaches are widely discussed in literature for the modeling of road irregularities:

1. Synthesis using harmonic functions and randomly distributed phases [12]
2. Synthesis on the basis of noise-excited linear dynamic systems [13,16]

Both methods can be used for single or multitrack model processes; the achievable model quality is equivalent in principle. Linear filters seem more practical, however, for online generation as they

minimize the data management and require less computational input. A road excitation process with waviness $w = 2$ can already be generated by means of a simple first-order filter, for example:

$$\dot{\xi} + \omega_g \xi = \sqrt{v S_0 \Omega_0^2}\, z; \quad z \equiv z(k\Delta t) = \sqrt{\frac{12}{\Delta t}} Z_k; \quad k = 0,1,2,\ldots \tag{28.12}$$

whereby Z_k is a random number sequence in the range of [−0,5;0,5] which is called up in sufficiently small time intervals Δt. The road excitation ξ approximates the desired target spectrum for frequencies above $f \succ f_g = \omega_g / 2\pi$ with intensity S_0 and waviness $w = 2$. For more refined approaches toward filter approximation of road spectra, refer, for example, to [12,14,19].

28.2.5 Synthesis of Two-Track Excitations

Two-track road excitations with coherence characteristics close to reality are generated through a straightforward mixing function. Two independent excitations ξ_1, ξ_2 are used as input signals which are already adjusted to the desired longitudinal direction spectrum. From these signals, two coupling signals η_1, η_2 are generated through first-order filters:

$$\eta_1' + \Omega_{\Delta y} \eta_1 = \Omega_{\Delta y} \xi_1; \quad \eta_2' + \Omega_{\Delta y} \eta_2 = \Omega_{\Delta y} \xi_2 \tag{28.13}$$

Thereby and with the original excitations ξ_1, ξ_2, the desired road excitations ξ_l, ξ_r are generated:

$$\xi_l = \xi_1 - \left(1 - \frac{\sqrt{2}}{2}\right)\eta_1 + \frac{\sqrt{2}\eta_2}{2}; \quad \xi_r = \frac{\sqrt{2}\eta_1}{2} + \xi_2 - \left(1 - \frac{\sqrt{2}}{2}\right)\eta_2 \tag{28.14}$$

The synthesis maintains the longitudinal spectra and generates for the cross-spectrum/coherence:

$$S_{lr} = S_{rl} = \gamma S_{\xi\xi} = \frac{\Omega_\gamma^2}{\Omega^2 + \Omega_\gamma^2} S_{\xi\xi} \tag{28.15}$$

and therefore a characteristic familiar for road coherences ($w_\gamma = 2$).

28.3 TIRE MODELS

The development of tire models for vehicle dynamics and ride comfort analyses has progressed significantly over the last few years, both with regard to the model quality and the introduction of quasi standards. Three major model classes with differing complexity and application spectra have now become evident [28]:

1. Map-type models (Pacejka-type models and others) [22]
2. Part-physical models, brush and ring models (e.g., BRIT, MF-SWIFT [32])
3. Detailed tire structure models (e.g., FTyre, RMOD-K [29,30]) (Figure 28.10)

Tire models more or less completely include the elements required for tire analysis:

- Steady-state tire force and torque characteristics
- Vertical stiffness (and damping) properties

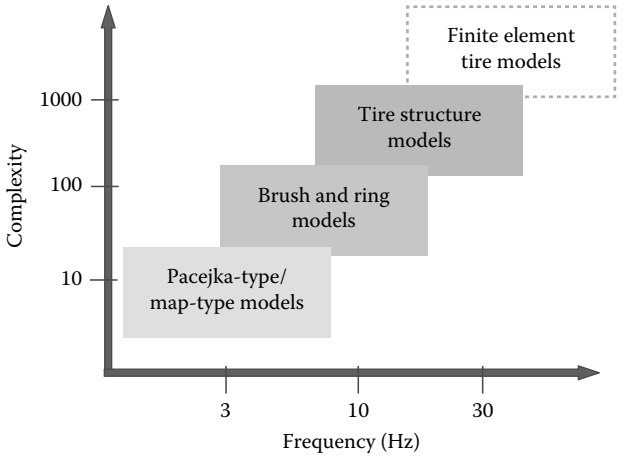


FIGURE 28.10 Rough classification of tire model designs.

- 3D-kinematics analysis/calculation
- Road surface contact evaluation
- Contact patch kinematics and kinetics modeling
- Tire dynamics/vibrations modeling concepts

28.3.1 Application Scope and Related Demands/Requirements

The approved operating range of the different model types overlap nowadays due to constant improvements. Basically, map-type models are primarily suited to typical vehicle dynamics and handling analyses. Part-physical models are required when the database is weaker, but also where severe control system interventions—for example, ABS or ESP interventions—form the focus of the analysis. For comfort and/or vibration analysis, and also for the calculation of durability load profiles, on the other hand, detailed tire structure models must be used in order to attain reliable results. From the current perspective, this model class is still in its industrialization phase. However, practice-related validations against test drive results have not yet been completed. Figure 28.11 gives an overview of the application scope of different tire models [27].

28.3.2 Road, Tire, and Vehicle Interfaces

In vehicle dynamics and driving comfort simulation, tire models are normally formulated as generalized force elements. The input values are the position $\vec{r}_c$, angle $\vec{\varphi}_c$, and speed variables $\vec{v}_c$, $\vec{\omega}_c$ of the rim center imposed by the vehicle. The tire provides feedback from the resulting tire forces $\vec{F}_c$ and torques $\vec{M}_c$, as illustrated in Figure 28.12. During this process, the internal tire contact conditions to the road surface are determined and evaluated. For this purpose, the tire model requires additional information about the road surface contour and, if required, the traction conditions.

28.3.3 Tire Model Parameterization

The parameterization and validation of tire models generally pose a number of problems, as both driving measurements and sophisticated test rig measurements are associated with insecurities and errors. In the case of road measurements, for example, the local friction values cannot be documented reliably, the lateral speed data are relatively imprecise, and the resolution of the force and torque measurements leaves a lot to be desired. Test rig measurements are normally better with regard to the aforementioned issues due to the laboratory environment. However, tire test machines

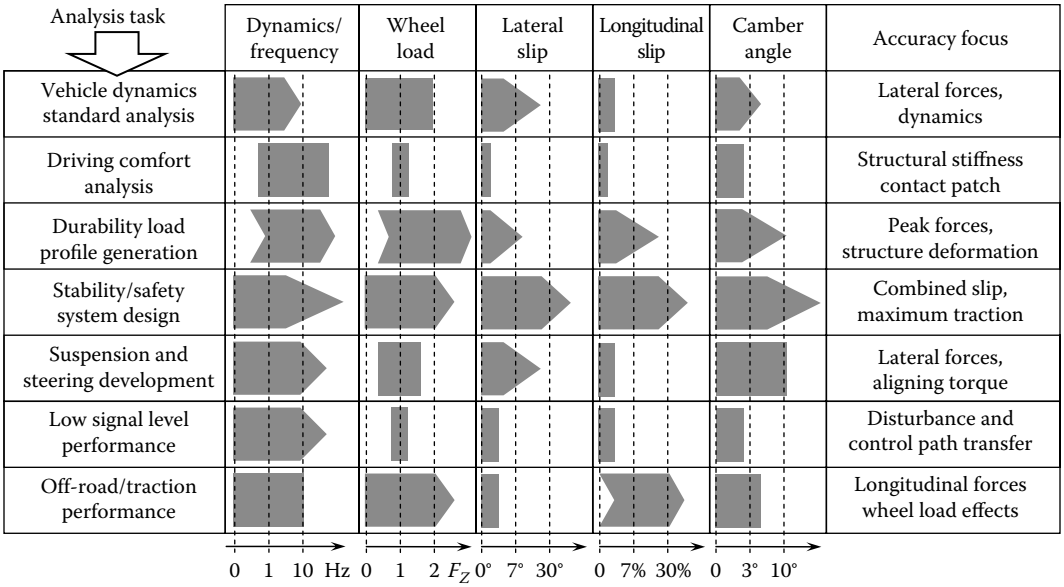


FIGURE 28.11 Application scopes and requirements for tire models for vehicle dynamics, driving comfort analyses, and for durability load profile generation.

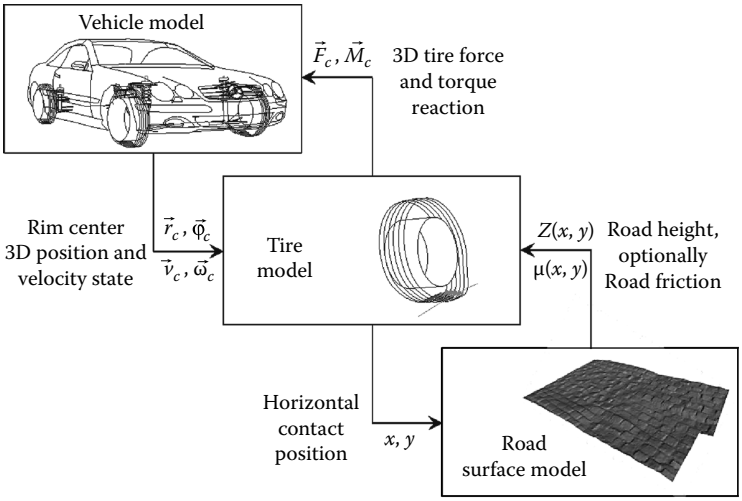


FIGURE 28.12 Vehicle–tire interface and tire–road interaction (refer to Sec. R).

always involve compromises with regard to the road contour (drum curvature, external or internal drum), the road surface (synthetic linings on flat surface and external drum test rigs), and the thermal stressing of the tires. In addition, test rigs are not normally capable of simulating dynamic load cases (exception: road disturbances such as cleats or vertical steps on internal drum test rigs).

Due to these measuring technology restrictions, some tire properties are very well known, others only sketchily. Steady-state tire characteristics have been relatively well researched and can be measured reliably for different wheel loads F_z and different camber angles γ:

- Lateral force side slip characteristics $F_q(\alpha;F_z)$
- Aligning torques side slip characteristics $M_R(\alpha;F_z)$
- Longitudinal force slip characteristics $F_l(s;F_z)$

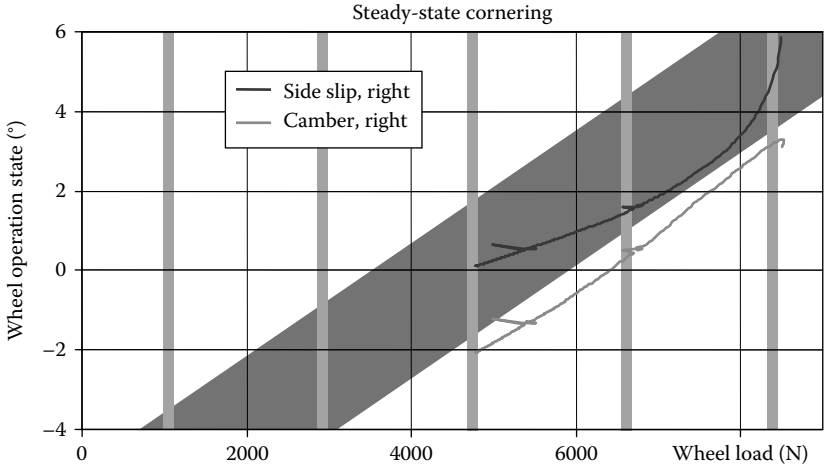


FIGURE 28.13 Comparison of the measuring ranges between conventional tire measurement procedures at constant wheel load (light grey, vertical) and the TIME measuring procedure (dark grey), which is focused on the slip angle–camber–wheel load combinations achievable in dynamically steady state (e.g., steady-state cornering).

This also applies to some extent to combined load cases, for example, $F_q(\alpha;F_z)$ for fixed longitudinal slip s or $F_l(s;F_z)$ for fixed slip angle α.

Measurements at constant wheel load F_z support our understanding of the tire behavior and are well suited to tire model parameterization. Constant wheel loads are not so interesting during driving operation, however. This is why the TIME group has suggested an alternative, vehicle-dynamics-related measuring procedure. The TIME procedure focuses on typical cornering maneuvers, where side slip and camber angle variations are correlated, as illustrated in Figure 28.13.

The TIME concept has two central objectives:

1. To raise as many measuring values as possible in operating points that are dynamically relevant and achievable
2. To protect the measuring tire from high thermal loads, that is, reduction of the measuring range itself to a minimum, reduction of the vehicle speed during measurement, as the related faults are smaller than the effects of excessive tire stress

In the meantime, Pacejka model approaches specifically matching this measuring scheme have been suggested for the approximation of the tire force characteristics [23]. It would be desirable for approaches of this type to establish themselves as standards. However, even in this case, any wider use in the future must demonstrate whether the restriction of the measuring data to essential vehicle dynamics operating points is adequate to determine the tire behavior with sufficient completeness.

28.3.4 Tire Dynamics and Low-Speed Properties

In order to simulate the dynamics of the lateral and longitudinal force generation, the map-type models must (at least) be expanded. In the case of a known lateral tire stiffness c_q^R, this can be realized as follows, for example, for the lateral force dynamics:

$$\frac{\dot{F}_q}{c_q^R|v|} + F_{q,stat}^{-1}(F_q) = \alpha; \quad F_{q,stat}(\alpha) = \mu_{\max} F_z^{eff} f\left(\frac{\alpha}{\alpha_{\max}\mu_{\max}}\right) \tag{28.16}$$

This physical lateral force dynamics approach is based on the fact that the tire road contact patch is coupled via the lateral tire stiffness c_q^R to the rim. Herein, the steady-state tire force $F_{q,stat}$ is defined by a simple map-type approach, based on the friction coefficient μ_{max}, the vertical force F_z^{eff}, a normalized tire force shape function $f(\ldots)$, and the side slip angle α_{max} at side force maximum.

For small displacements, the approach is equivalent to the established relaxation length concept. In contrast, type ($T1$) results—physically consistently—in faster contact force changes for higher slip angles. In addition, the wheel load influence on the lateral and longitudinal force dynamics is represented consistently: This is a typical parameter excitation in the denominator of the first-order system. This explains not only the fast lateral force response to wheel load reductions but also the delayed lateral force build-up in case of wheel load increases. In the same way, the longitudinal force dynamics can be modeled as well. (It is expedient to use steady-state radial forces for the unstable part of the radial force characteristic in order to avoid model-induced limit cycles.)

Another problematic field of tire-modeling technology is the operation at low speeds or at standstill. Phenomenological models generate infinite stiffness values during standstill as the crucial input values—longitudinal slip and side slip—explode as a result of the $|v|$ division. The models are prepared for operation at low speeds by means of more or less elegant numerical tricks such as restrictions of the minimum speed. (This is not a problem for physical models as they use the actual material and structural properties.) The quality of the $v \Rightarrow 0$ implementation can easily be tested by simulating stopping and restarting on a slope, for example, or by steering with different steering rates during standstill. The latter should always yield the same turnslip torques at small steering rates. At the beginning and when the steering direction is changed, the (approximately) constant stiffness of the belt–rim coupling should become visible.

28.3.5 Friction Behavior and Modeling

Tire analyses are most frequently carried out on dry, high-grip road surfaces. Some in-depth studies are available for moist and wet roads [24], and the same applies to tire behavior on snow and ice. However, it can be assumed that the properties of tires have improved in the meantime compared with what is suggested by some studies which are a few years old [25].

Most tire models now allow the overall tire-to-road traction to be specified, that is, to be parameterized. This option allows, for example, to adjust the tire model traction characteristics to specific test track properties.

Looking at part-physical tire models or tire structure models based on the brush model principle, it is necessary to know the actual friction conditions between tire rubber and road in order to be able to parameterize the models sensibly.

The sliding friction properties of the rubber are a small problem in this context, as it has been shown within the framework of the TIME analysis that the vehicle speed has little influence on the tire characteristics (i.e., the sliding speeds attainable in the tire contact area are comparatively small). The sliding friction properties can be regarded as almost constant within the framework of the relevant speed range (exception: driving stability test procedures). They can therefore be reliably identified from the global tire characteristics.

The determination of the (material) friction conditions between tire rubber and road raises more problems on the other hand. Due to the strong coupling with the ground pressure distribution in the model, pure identification approaches are prone to errors. In order to avoid this dilemma, individual direct measurements of the rubber–road friction have been taken on surfaces similar to roads [26], as illustrated in Figure 28.14. These show that the static friction coefficient—in line with the brush model concept—can reach significantly higher values than the global tire–road friction. It also becomes clear that the local ground pressure and therefore the homogeneity of the ground pressure distribution significantly influences the road adhesion potential of the tire.

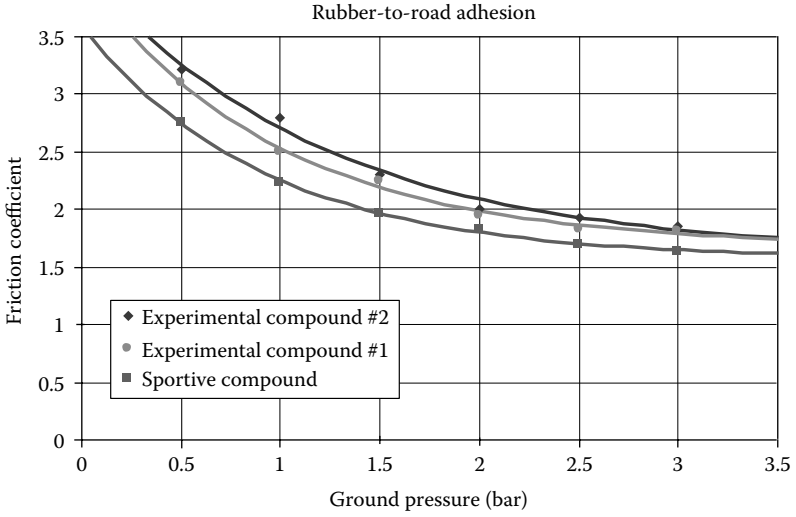


FIGURE 28.14 Material friction capability of a ground-in tire rubber rib on a high-grip road surface (experimental rubber compounds and sportive compound).

28.3.6 Tire Comfort Behavior and Modeling

The tire comfort properties are usually discussed by analyzing the disturbance transfer behavior. In this context, the vertical tire spring rate and the tire dynamics while driving across cleats or—closer to practice—vertical steps are crucial. Figure 28.15 shows such a test rig drive across vertical steps at 80 km/h under side slip stress (equivalent to 30% of the maximum lateral force).

The force and torque curves clearly show the belt vibrations of the tire at different frequencies, from 35 to 350 Hz. The tire vibrations only decrease slowly. This results in the real adverse effect on comfort, as these road-induced vibrations excite other vehicle components in the chassis and body so that it takes some time (despite the higher-frequency excitation) until the whole system calms down again.

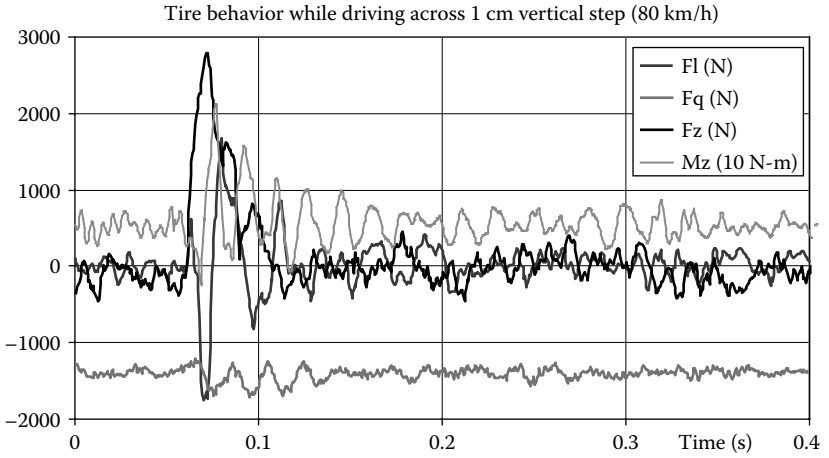


FIGURE 28.15 1 cm–Standard passenger car tire driven at 0.06 s across a vertical step at 80 km/h; for obstacle type, refer to Figure 28.16.

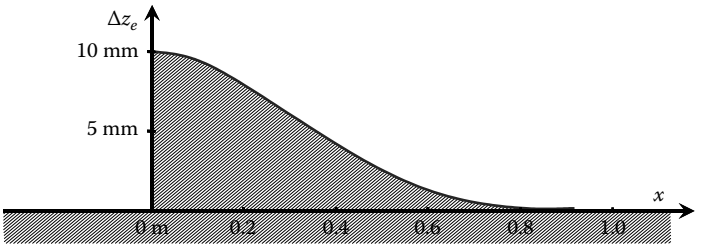


FIGURE 28.16 1 cm–Vertical excitation (+1 cm spontaneous increase, then drop to zero according to second-order system with axle frequency adjustment).

In principle, tire structure–related models such as FTyre or RMOD-K are capable of describing phenomena of this type with sufficient precision. They therefore form the basis for systematic optimization of the tire–chassis–body comfort (Figure 28.16).

Similarly, driving situations for durability analyses should become capable of simulation. It must be taken into account, however, that the force and excitation amplitudes are higher by more than a factor of ten in these cases. Relevant test cases for test rig analyses and validations are currently being examined. It can be assumed that the conventional tire models must be expanded in some form or another in order to supply correct results even under extreme loads. (As the durability of a component typically degenerates with the fourth power of the force amplitude, such simulations need to be extremely precise.)

28.3.7 Challenges in Future Tire Modeling

Some tire properties and phenomena will probably be difficult or impossible to integrate into vehicle/axle dynamics tire models or remain restricted to purely qualitative estimates. This applies in particular to the thermal conditioning of the tire while driving and the related changes in the coefficient of friction, as the necessary flow information is not normally available [37]. This similarly applies to certain comfort effects such as the reduction in the radial stiffness during rolling as against tests during standstill (caused by micro-distortions of the contact patch against the road surface), thermally induced settling effects such as flat spotting, or more generally the prediction of tire comfort characteristics at different operating temperatures.

Instead, the focus will continue to lie on the more precise modeling of the tire at one or a few selected reference temperatures. A measuring program for the largely comprehensive documentation of the tire properties relevant to chassis design requires comparatively elaborate measurements even under these restrictions [27]:

1. Steady-state properties
   a. Steady-state tire force and torque characteristics
   b. Vertical stiffness and damping properties
   c. Standstill aligning torque (turnslip operations)
2. Dynamic properties
   a. Lateral pull experiments (lateral stiffness)
   b. Step steer excitations
3. Vibration and high load properties
   a. Cleat excitations and/or vertical step excitations
   b. Durability load obstacles
   c. Structural stiffness measurements

Another modeling field in the next few years is likely to be a more precise simulation of run-flat tires, as these systems, on the one hand, require the understanding of tire behavior at extremely low

air pressures and, on the other hand, increasingly transfer the load-carrying function from the traditional air suspension-type concept toward a complex rubber-fiber compound structure.

28.4 AXLES: CHASSIS COMPONENTS

The wheel suspension ensures that the translational and rotary displacements of the wheel/road contact associated with jounce and steering motions do not as a whole impair the vehicle behavior, or have a positive influence. This is normally achieved by adjusting the basic kinematics such that vertical road disturbances only exert a small influence on the wheel position and therefore on the force transmission. In analogy, the centers of rotation crucial for the drive and brake-force transmission are selected such that the related pitch and lift motions are largely suppressed.

Besides the kinematics, chassis systems make varied use of the elasticity and damping of chassis mounts and axle components in order to

- Optimize vehicle response in critical driving situations
- Improve ride comfort
- Shape and adjust basic vehicle characteristics such as straight-line stability, precise steering response, crosswind response, etc. in a targeted way

As an example, Figure 28.17 shows the front and rear suspension of a typical mid-class passenger car.

The perceivable ride characteristics of a vehicle are therefore increasingly more dependent on the properties of plastic and rubber materials than conventional metals.

28.4.1 Chassis Mounts and Bearings

In general, three different types of connection elements are typically used as chassis mounts:

- Rubber bushings
- Hydro mounts (in particular for the connection of subframe and body)
- Ball joints

The dynamic properties of rubber mounts and of hydro mounts cannot sufficiently be described in terms of mechanical base elements such as springs, viscous damping elements, or masses. The main reason for this is the complex material properties of rubber products. Under quasi-static load, rubber

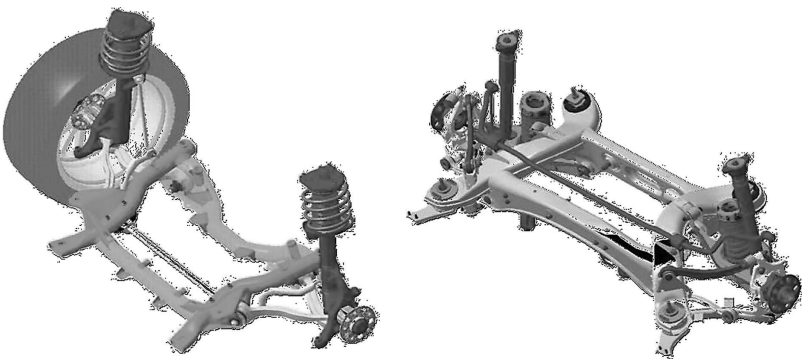


FIGURE 28.17 Wheel suspension of a typical compact model inclusive of subframe and suspension system (front: spring-and-damper wheel location; rear: multilink rear suspension).

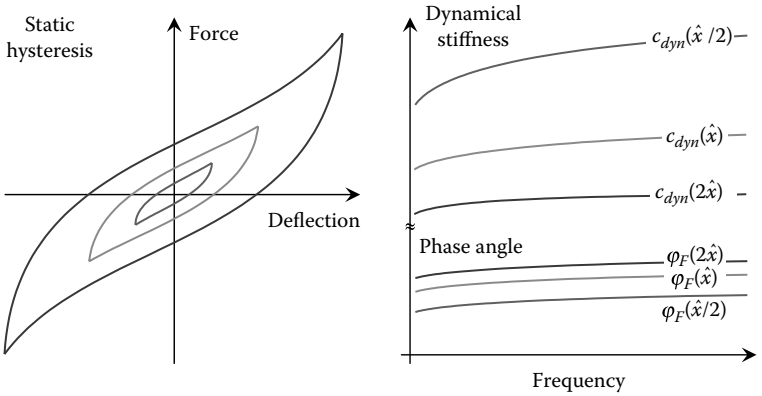


FIGURE 28.18 Quasi-static and dynamic behavior of a typical rubber mount component.

components show pronounced hysteresis effects, and under dynamic load they stiffen, as illustrated in Figure 28.18. The phase and loss angles are mostly small, however, and only increase slightly with the frequency.

In order to model the nonlinear component properties of rubber mounts, multicomponent models are often used which synthesize the observed effects by means of appropriately configured elements. In this process, a nonlinear spring element simulates the basic stiffness, as illustrated in Figure 28.19. The hysteresis effects can be modeled by means of a type of hysteresis generator, that is, a force element explicitly generating the typical curve shapes depending on the preceding movements. Besides these static approaches, Maxwell elements and, if required, a normal damping element are added in order to model the dynamic stiffening under increasing frequency.

The modeling of a rubber mount dynamics with Maxwell and damping elements proves problematic in practice as, with increasing approximation quality, the risk of generating negative loss angles at high frequencies increases. This type can be avoided by means of appropriate restrictions; their very existence shows, however, that shape functions do not ideally suit the system properties.

To avoid these difficulties in principle, an alternative concept has recently been suggested for modeling rubber stiffening on the basis of fractional derivatives [42]. The core idea is that almost all constant phase angles and the amplitudes of rubber hardening, which increase degressively with the frequency, correspond to a dynamic behavior that lies somewhere between the spring and damper. Fractional derivatives allow us to formulate such characteristics.

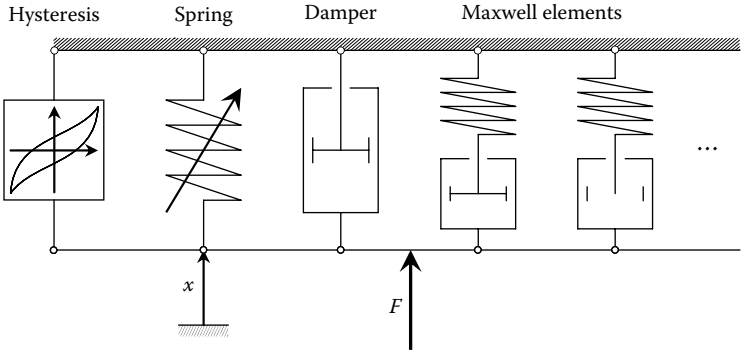


FIGURE 28.19 Synthesis model for a chassis rubber mount.

The fractional derivative $D^{\alpha}\{x(t)\}$ is defined in analogy to the known integer derivative of a time function $x(t)$:

$$y(t) = D^{\alpha}\left\{x(t)\right\} \equiv \overset{(\alpha)}{x}(t) = \frac{d^{\alpha}}{dt^{\alpha}}x(t); \quad \alpha \in \Re \tag{28.17}$$

The signal behavior of the α-times-differentiated signal $y(t)$ therefore results in the frequency domain from the following expression:

$$Y(\omega) = (j\omega)^{\alpha}X(\omega) = \omega^{\alpha}\left(\cos\frac{\alpha\pi}{2} + j\sin\frac{\alpha\pi}{2}\right)X(\omega) \tag{28.18}$$

The phase or loss angle is therefore constant at $\alpha\pi/2$. For signal x available at discrete time steps $\Delta t(x_k = x(k\Delta t))$, the calculation of the fractional derivative y can be reduced to a difference quotient evaluation:

$$y_k = D^{\alpha}\left\{x(t)\right\} = \lim_{n\to\infty}\sum_{l=0}^{n-1} a_l x_{k-l}; \quad a_l = \frac{\Delta t^{-\alpha}}{\Gamma(-\alpha)}\frac{\Gamma(l-\alpha)}{\Gamma(l+1)} \tag{28.19}$$

Herein, $\Gamma(n+1) = n!$ is the so-called gamma function, a continuous expansion of the factorial. In practice, the series expansion must be truncated at a predefined ordinal number n. This results in certain deviations from the continuous approach (28.18) at low frequencies, but should be acceptable for most applications as the coefficients a_l converge with increasing l quickly toward zero.

The frequency response of the discrete fractional model system finally results as follows:

$$F^{n}_{x>y}(\omega) = \frac{Y}{X} = \sum_{k=0}^{n-1}(\cos k\omega\Delta t - j\sin k\omega\Delta t)a_k \tag{28.20}$$

Figure 28.20 shows the transfer behavior of a standardized discrete fractional model (time step width $\Delta t = 3$ ms, system orders $n = 10$ and $n = 100$) in comparison to an exact, continuous approach at a derivative step of $\alpha = 0.15$. This corresponds to a loss angle of 13.5°. This shows clearly that this approach fits in well with the rubber-typical, almost-constant phase responses and with the amplitudes which increase degressively with increasing frequency. The deviations from the continuous approach at small frequencies are essentially determined by the absolute filter length, that is, product $n\Delta t$. The differences are insignificant, however, as only the deviations from the real mount frequency responses are decisive for the quality of the mount modeling.

In addition, there is a series of longitudinal rotary bearings, for example, in shock absorber struts or shock absorbers, which normally result in deflection-dependent bending stiffness values and in quite complex load–friction interactions [4,43,47]. This phenomenon cannot be discussed in detailed here, however.

In general, the elastokinematic design of a chassis is based on the fact that the geometrical position of the wheel (location and angle) changes under the influence of the wheel forces and can be adjusted in a targeted way. Effects of this type can be used, for example, to improve the vehicle behavior while cornering, as illustrated in Figure 28.21.

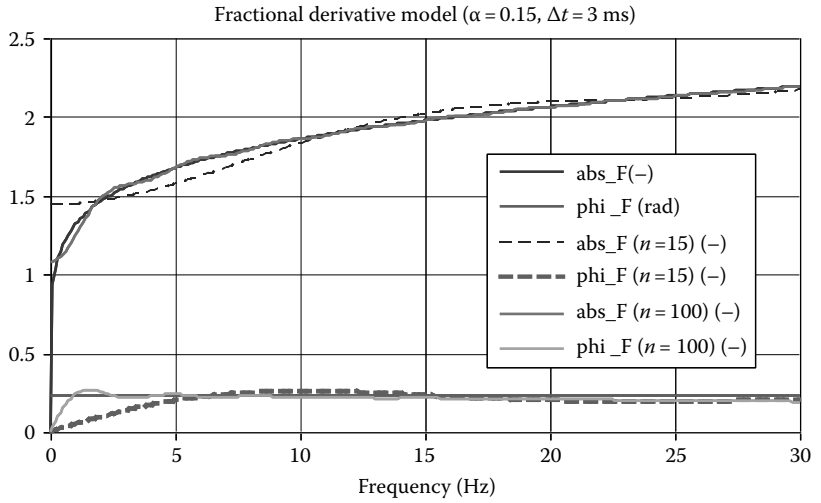


FIGURE 28.20 Frequency response of a rubber mount model on the basis of fractional derivatives.

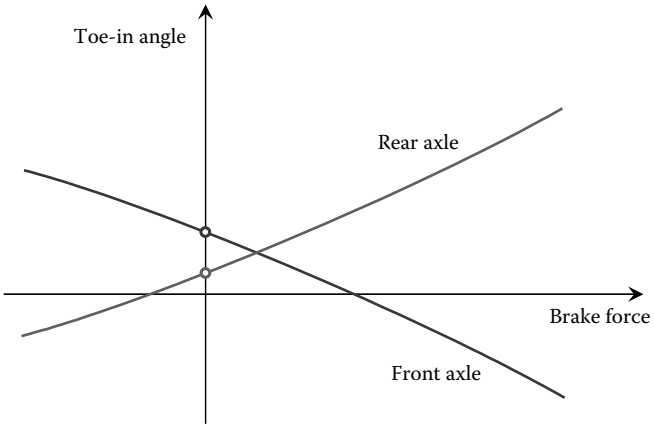


FIGURE 28.21 Elastokinematic optimization of the vehicle steering tendency during braking while cornering.

28.4.2 Elastokinematic Adjustment

Vehicle braking leads to displacement of the wheel loads toward the front. Consequently, the front wheels generate somewhat higher lateral forces—at a given slip angle. The lateral forces at the rear axle decrease accordingly. This is further reinforced by the additional stress due to the longitudinal force. This would result in the vehicle turning strongly toward the inside of the corner.

To avoid irritations for the driver, attempts are made to ensure that the vehicle behavior changes only a little or not at all during braking. This can be achieved by means, among other, of the measures shown in Figure 28.21. Here, the front axle goes into toe-out during braking, the crucial outer wheel therefore reduces its slip angle, and the disturbing increase in the lateral force is avoided. Vice versa, an additional side slip angle is generated elastokinematically at the rear axle counteracting the wheel-load-induced reduction in the lateral force or compensating it. Altogether, the resulting vehicle behavior is largely neutral to brake applications while cornering. (In contrast, the toe-in changes during μ-split braking should be adjusted in the opposite direction in order to improve vehicle directional stability.)

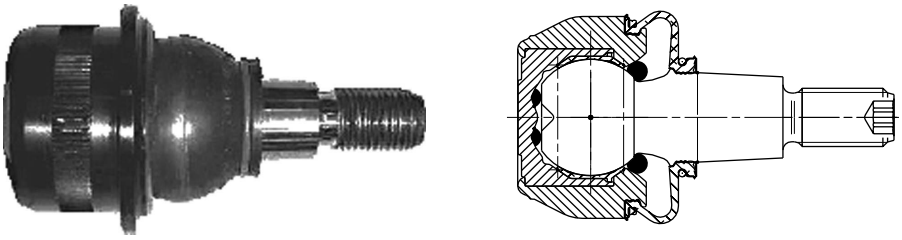


FIGURE 28.22 Chassis ball joint (cross section).

As the force transfer mechanisms are generally different for drive and brake forces and steering torques in the axle, noticeable improvements in the vehicle behavior can be achieved at least in the case of multilink suspensions, but within certain limits also for simpler axle concepts, by means of targeted force control interventions. The functionally relevant components for the elastokinematic design are the chassis mounts and—to a certain extent—the larger elasticities of the axle components themselves. In addition, chassis mounts fulfill a number of decoupling and isolating functions as they provide for a more resilient wheel–vehicle connection and therefore help suppress the transmission of disturbing excitations. However, synthetically introduced elasticity always promotes the restraint and natural vibration potential of the chassis. During chassis adjustment, a sustainable compromise must therefore be worked out between global control and isolating properties and the self-excitation tendencies of the chassis.

From a modeling point of view, probably the most simple and clear chassis mounts are ball joints, typically installed close to the wheel (refer to Figure 28.22). These joints generally demonstrate a comparable stiffness in all directions in space. The stiffness values reach around 10 kN/mm or more. Compared to other axle components, these are therefore quasi-rigid connections. The damping effect is correspondingly small or even negligible. Due to lifetime requirements, ball joints must be strongly pretensioned internally. This means that rotating movements are always associated with larger friction torques of some newton meters. To be able to correctly simulate the influence of the mount phenomenologically, especially with clearly differentiated adhesion and sliding states, some special integrator settings must be determined even for modern MBS systems.

28.5 SUSPENSION SYSTEMS

Suspension systems in vehicles have a number of different functions; they

- Carry the load of the vehicle body
- Compensate road irregularities and torsion effects
- Dampen and smooth road and vehicle dynamics–induced vibrations
- Ensure that the wheel travel limitations are sufficient

so that these functions can be fulfilled in practically all possible driving situations without damage to the vehicle or chassis.

Depending on the chassis design, spring–damper systems are either focused on separate components such as in multilink or double wishbone axles, or integrated into the wheel suspension system, for example, in McPherson or shock absorber strut independent suspensions. The latter generally results in a reduced installation space and weight, although it usually requires more elaborate detail solutions in order to suppress the mutual interferences of the design elements in a restricted space, for example, sophisticated strut top mounts that separate the vertical spring, damper, and lateral force transfer paths.

28.5.1 Steel-Spring Suspension and Suspension Characteristics

With regard to modeling, the body springs, that is, the supporting springs and stabilizer bars, are comparatively straightforward components. For their simulation with good approximation, it is sufficient to know the partly nonlinear spring characteristic. For individual wheel suspensions, the wheel jounce or distance of the spring mount points form the model input value for the suspension model; in the case of rigid axles, on the other hand, both wheel jounce values of an axle contribute to the spring deflection. For the stabilizer bar displacement, the difference in the jounce values between the left-hand and right-hand vehicle side is decisive; they generate a counteracting force couple that primarily counteracts roll movements.

For more precise analyses of the wheel suspension or axle dynamics, it is sometimes necessary to explicitly model the kinematics and design details of the spring and/or stabilizer bar construction and their mounts. For example, the internal vibration modes of a coil spring play a role with regard to the tire comfort. In the same way, the stabilizer bar mounting and, in the case of complex stabilizer shapes, maybe also their force transfer perpendicular to the actual effective direction, can improve, but also impair, the wheel suspension and steering properties.

The contribution of suspension spring components to the wheel suspension movements is especially obvious in the case of leaf springs guiding rigid axles. They are often used in commercial vehicles, in particular, because of their extraordinary robustness. The calculation of the kinematics and elastokinematics is particularly time consuming in this case, as many effects that can otherwise be separated are strongly coupled. An example is the support of the brake forces by the leaf springs, which results in axle pitching and jounce effects during braking and, therefore, among others, couples the brake control adjustment with the chassis design.

28.5.2 Air Suspension and Air Suspension Modeling

In contrast to steel spring suspensions, air suspensions and hydropneumatic arrangements generally require more detailed consideration of the relevant media and interconnections. In both cases, the actual suspension effect is achieved by means of compression of a gaseous medium. In the case of air springs, the operating pressure level usually amounts to some 10 bar, while hydropneumatic systems commonly reach operating pressures of 50–80 bar. Since typically simple, low-molecular gases are used, the pressure changes can be determined on the basis of the ideal gas equations. For slow, isothermal processes and fast, adiabatic volume changes $\dot{V}$, the edge case models apply (adiabatic exponent κ):

$$\dot{p}_{isotherm} = -\frac{p}{V}\dot{V}; \quad \dot{p}_{isotrop} = -\kappa\frac{p}{V}\dot{V} \tag{28.21}$$

In practice, the system always exchanges thermal energy with the environment. This results in a delayed pressure change:

$$\dot{p} + \omega_g p = -\kappa\frac{p}{V}\dot{V} + \omega_g\frac{p_0 V_0}{V}; \quad p_0 = p(T_u); \quad V_0 = V(T_u) \tag{28.22}$$

The heat transfer effects of the system are summarized in the limit frequency constant ω_g, p_0 represents the system reference pressure at ambient temperature T_u, and V_0 is the corresponding system volume. Depending on the system size and surface design, limit frequencies ω_g indicating delay times up to several seconds can occur—the latter predominantly in commercial vehicle air suspension systems. (This makes the technical analysis and assessment of air suspension systems more difficult, especially on laboratory rigs, as the thermal conditions must be precisely determined and checked.)

Besides comfort improvements mainly based on very low spring rates, gas-operated suspensions also have the advantage that the spring rate adjusts itself to the wheel load and therefore the vehicle load. The reduction in the natural system frequency with increasing load, which can normally be observed, is therefore avoided and in some cases even overcompensated.

Air suspension chassis also allow the spring rates to be adjusted to the vehicle operating requirements—by means of an appropriate design of the bellows characteristics. Similar results can be attained in the case of hydropneumatic suspensions by using several accumulators, whereby design possibilities are quite limited. Hydropneumatic chassis are of greater interest as a basis for active suspensions as they can be installed in a small space at reasonable cost and as they can easily be activated. At present, some tire comfort compromises due to problematic high-pressure sealing properties still exist which need to be overcome.

28.5.3 Dampers and Damper Modeling

Vehicle dampers are more complex with regard to modeling, as due to their design they respond very differently, depending on stress amplitude and frequency. Figure 28.23 shows the design of a single-tube damper that represents the most simple damper design. The nonlinear restrictor valves $\Delta p(s)$ installed in the damper piston determine the actual damper characteristic $F_d(\dot{x})$:

$$F_d = A_k p_k - A_s p_s; \quad p_k - p_s = \begin{cases} \Delta p_{Druck}(s) & \text{für } \dot{x} \geq 0, \\ \Delta p_{Zug}(s) & \text{für } \dot{x} \leq 0; \end{cases} \quad s = (A_k - A_s)\dot{x} \tag{28.23}$$

In order to be able to absorb the volume of the damper piston rod during compression, a small pressure chamber V_G is arranged on the damper floor. The gas contained therein is precompressed, usually at pressures p_G of 10–30 bar, so as to keep the rod side at positive pressures at all times. The precompression pressure is associated with a damper base force that is always positive:

$$F_G = (A_k - A_s)p_G \tag{28.24}$$

Due to the coupling with the gas volume, the piston pressure $p_k \cong p_G$ roughly corresponds to the gas pressure. The piston pressure fluctuations are relatively small, that is, the damper force spread occurs on the rod side.

As dampers are required to remain functional over a relatively long lifetime, high-performance seals and guides are used. In combination with the pressure preload, this means, however, that dampers will tend to "harden" at small excitation amplitudes. The force response of the damper then includes a pronounced elastic content; the phase turns from 90° into the direction of 0°.

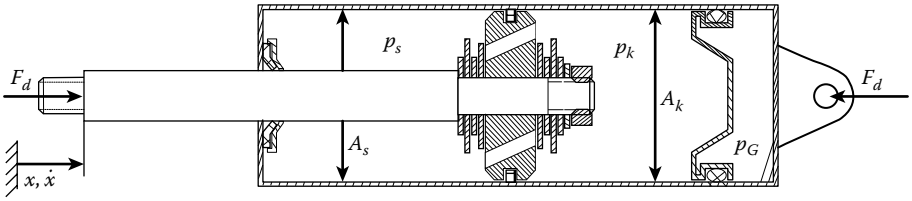


FIGURE 28.23 Design of a single-tube gas pressure shock absorber.

As the seals have a certain longitudinal elasticity (longitudinal stiffness c_D), the hardening effect can be modeled elegantly using spring-friction elements:

$$F_R = \min\left(\max\left(c_D \int \dot{x}dt;\quad -F_R^{\max}\right); F_R^{\max}\right) \tag{28.25}$$

Besides the external seal, the damper piston seal and the dividing piston contribute to the friction and hardening effects. In this process, the maximum friction force $F_R^{\max}$ is influenced by the current damper pressures p_s, p_k. In the event of higher-frequency damper stresses, the compressibility β of the damper medium has to be taken into account as well:

$$\beta = -\frac{\partial V}{V \partial p} \approx \frac{3 - 8 \times 10^{-5}}{\text{bar}}. \tag{28.26}$$

Typically, the effective compressibility β_{eff} is somewhat higher, as the damper body and sealing elements contribute additional elasticity. Examples of the consequences of the hardening and dynamic effects are shown in Figure 28.24.

In the event of very severe, fast damper excitations, the damper pressure on the rod side can briefly drop to steam pressure level. This results in local oil cavitation. Small steam bubbles emerge and collapse again. During normal driving operation, the damper design ensures that this operating state does not occur. During the so-called rough road testing, that is, during extreme durability tests on irregular tracks, it is hardly possible to avoid cavitation effects. To model the major effects, it is sufficient to adopt an additional pressure chamber at steam pressure level that fills dynamically when the steam pressure level is reached—and empties itself once the rod side pressure p_s increases again.

For the modeling of double-tube dampers, the same basic elements can be used again. However, the dividing piston is eliminated, and instead additional valve elements are to be added.

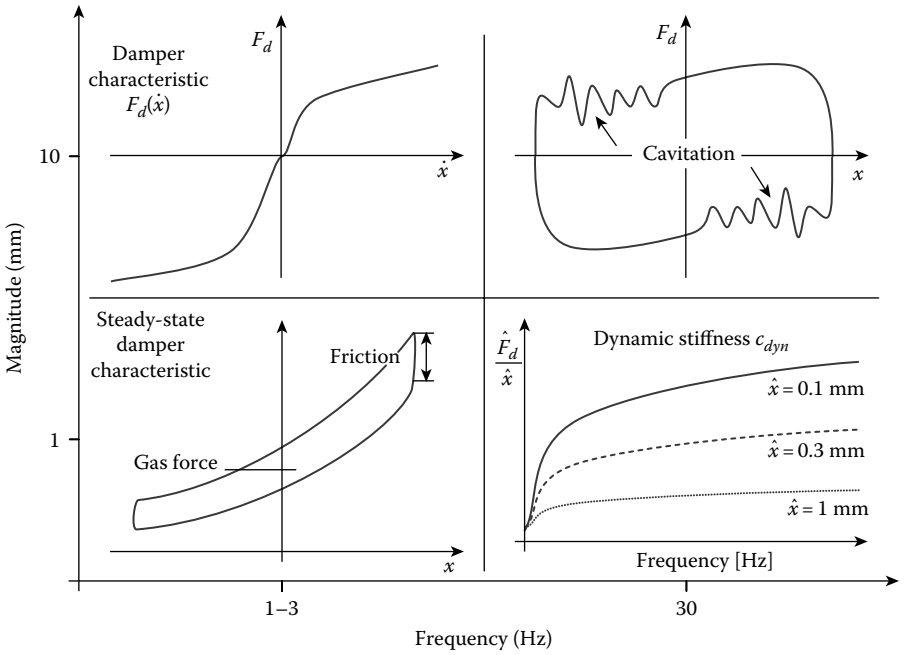


FIGURE 28.24 Aspects of damper characteristics in relevant operating ranges.

28.5.4 Variable Damper Systems

In the case of variable damper systems, bypass channels are usually added in order to allow the exchange of media between the rod and piston side by means of a switchable or continuously adjustable control valve. The basic damping is then normally increased strongly; control interventions therefore result in a reduction in the damping forces.

The related damper controls are mostly intended to adjust the damper hardness to the relevant driving situation. Fast control concepts influencing individual wheel movements have not so far been implemented in production vehicles, as on the one hand the required cycle rates result in high requirements with regard to computer capacity and signal processing, and on the other hand also damper speed measurements with very fine resolution would be necessary in order to make targeted use of the control intervention via the damping constant. Variable damping concepts influencing the force or differential pressure would resolve the sensor problem, but has not been so far realized robustly with the present technology.

Due to the inevitable hardening effects, dampers are normally connected to the chassis using rubber mounts. Their characteristics are adjusted such that the mounts react very softly to small deflections and therefore mask the friction effects. In the event of larger damper forces, however, high mount stiffness values occur so that the damping effect can impact on the chassis without major losses and act to reduce the body energy. All in all, a very low damping of smaller wheel vibrations of some 1–2 mm is therefore tolerated which does not noticeably reduce the ride comfort.

For commercial vehicles and some passenger cars, piston position–dependent dampers are sometimes used in order to be able to adjust the damping intensity depending on the load or in order to ensure by intelligent interconnection of bypass channels that the damper fills a pressure buffer during wheel movement, thus keeping the vehicle at its target level.

28.5.5 Push and Pull Stops

Due to the centering and longitudinal guides already in existence, the wheel movement stops are often integrated into the shock absorbers. The rebound buffers are usually straightforward progressive steel springs which "catch" the wheel elastically on lifting from the road before the mechanical spring travel limiters are reached. In contrast, the jounce buffers must be capable of absorbing sudden load peaks between the wheel and vehicle body in a limited installation space without endangering the body structure. Polymer foam stops fulfill this function quite efficiently as they counteract the jounce movement with high, progressively increasing forces. During the final rebound, they reach a much lower force level, however. This means that the mount directly reduces a significant amount of the kinetic energy. In modeling, these hysteresis effects can be approximated in a similar way as with rubber mounts (refer to Section 28.4).

28.5.6 Active Suspension Systems

Besides the conventional suspension systems described, a series of active suspension concepts have been developed over the last few years which are intended to resolve or at least attenuate the classic design conflicts between comfort and safety-oriented vehicle adjustments [45]. In principle, a distinction can be drawn between systems intervening on axles and individual wheel concepts.

The former activate the stabilizer bar, that is, they allow the imposition of stabilizer torsions generated by actuators or the generation of synthetic stabilizer forces. The main benefit of these systems is to influence the roll movements of the vehicle in a targeted way. (This is also where the internationally used designation ARS—active roll stabilization system—originates.) In most cases, the roll movements are simply suppressed by actuator elements up to the performance limits of the systems. The result is a vehicle which feels more stable and easier to control. Driving is

made much easier as the driver is actually no longer required to correct the vehicle roll by using his or her own orientation. However, this also means that these systems can be used to shape or to shadow vehicle roll properties that can be seen as a major and impressive vehicle stability indicator for normal drivers.

Technical realizations of active stabilization systems are based either on angular or torque actuators which are integrated either into the stabilizer itself or into hydraulic cylinders instead of the otherwise customary pendulum supports for stabilizer-wheel movement coupling. The latter are technologically more elegant, but require a certain minimum installation space. The modeling of the actuators is very specific to the system. In connection with pendulum support actuators, in particular, hydraulic diagonal coupling of the axles has been under discussion, in particular, with the aim of improving off-road performance. In individual cases, it will therefore be necessary to create very detailed models of the hydraulics networks to be able to optimize the layout and to assess and optimize the always very complex dynamic effects due to the normally rather long hydraulic connection lines.

Active suspensions controlling individual wheels fully include ARS systems as far as the range of functions regarding control of roll movements and wheel loads is concerned. They also allow for both vehicle level and vehicle pitch motions to be influenced in a targeted way. Figure 28.25 shows the two major design types: active hydro-pneumatic (AHP) and active body control (ABC) system.

In the ABC system, a hydraulic cylinder is arranged in series with a stiffer suspension spring. This allows the suspension level to be changed quickly and in a targeted way, sometimes against the shock absorber forces. Due to the series arrangement of suspension spring and actuator, ABC systems require a comparatively high installation space. The suspension spring, on the other hand, ensures that the overall comfort can reach the level of sophisticated conventional spring–damper systems. Depending on the system design, control limit frequencies of 3–8 Hz are reached, that is, the body motions can be controlled fully and confidently.

Active hydropneumatic systems are, in contrast, much more straightforward and space saving. Instead of the conventional spring–damper unit, a simple hydraulic cylinder is used that does not require significantly more space than a shock absorber. The basic suspension is taken over by a gas pressure accumulator. Damper valves are integrated in the connecting line which is kept as short as possible. The system is activated—in analogy with ABC or ARS systems—by controllable valves in combination with a high-performance pressure supply. The only, but decisive, disadvantage of AHP systems is the fact that friction effects in the load-bearing hydraulic cylinder give rise to a loss of comfort, which so far has ruled out its wider use in sophisticated vehicles.

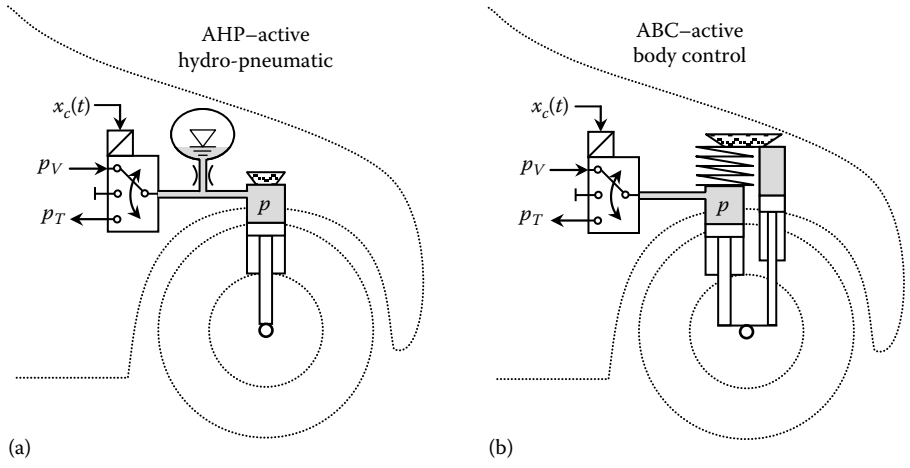


FIGURE 28.25 Types of active suspensions controlling individual wheels: AHP system (a) and ABC system (b).

The statements already made with regard to ARS systems equally apply to the modeling of active suspensions: Only on the basis of relatively elaborate detailed hydraulic system modeling can the dynamic transfer properties of the network be understood, interactions detected, and weak points or resonances be eliminated by targeted measures. For AHP and ABC systems, even more so than in the case of simpler ARS controls, the productive interaction between hydraulics, chassis mechanics, and control is decisive for good vehicle dynamics and comfort properties.

28.6 STEERING SYSTEMS

From the perspective of the driver, the steering system is supposedly the most important control intervention and also one of the most important sources of information about driving conditions.

28.6.1 Power Steering Systems

In order to keep the forces required to control the vehicle as small as possible, but at the same time allow sufficient design leeway for the optimization of the chassis in different velocity ranges, vehicles are today (almost) exclusively equipped with power steering systems. This means that the steering linkage, the connection between the steering wheel and axle steering, includes a reduction gear and at least one force-sensitive element as well as a hydraulic or electromechanical actuator.

Figure 28.26 shows a diagram of the most common steering design in use today, a hydraulic rack-and-pinion steering system. Figure 28.27 illustrates components of a vehicle power steering system. It becomes evident that many more than the basic elements described at the beginning are required for the design of a steering system.

Due to their high peak power, power steering systems require an efficient pressure supply. For this purpose, flow-controlled vane pumps are generally used with a delivery rate of up to 15 L/min and peak pressures of up to 150 bar. In smaller vehicles, it is possible to drive the power steering pumps by means of electric motors, thus avoiding a series of packaging and interaction problems.

Figure 28.28 illustrates the major hydraulic components of a vehicle power steering system. A torsion bar rotation, that is, the mechanical displacement of the torque-measuring element effects

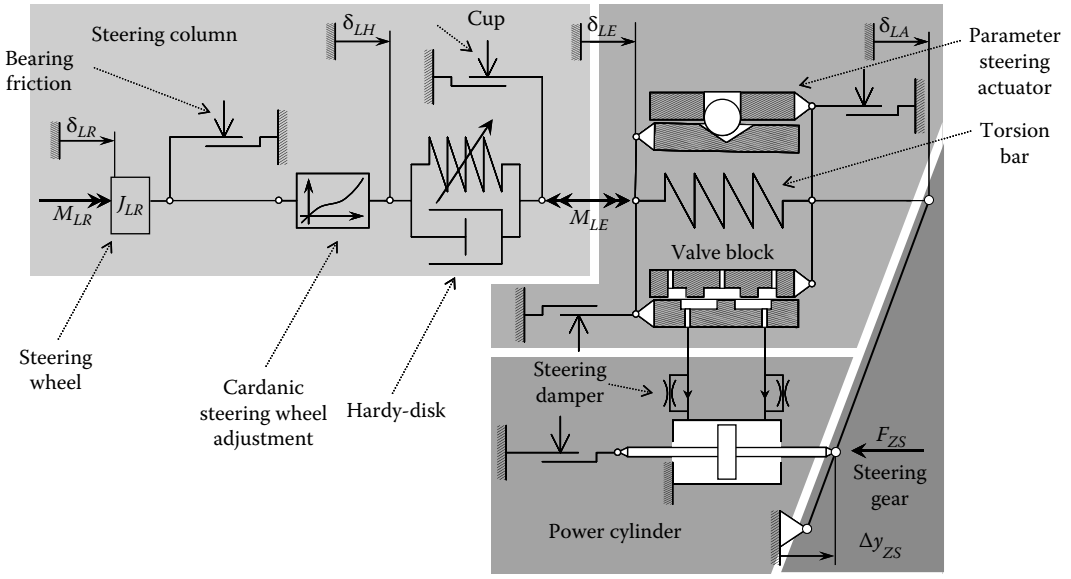


FIGURE 28.26 Mechanical function diagram of a vehicle power steering system, in analogy with a servo-hydraulic rack-and-pinion steering system.

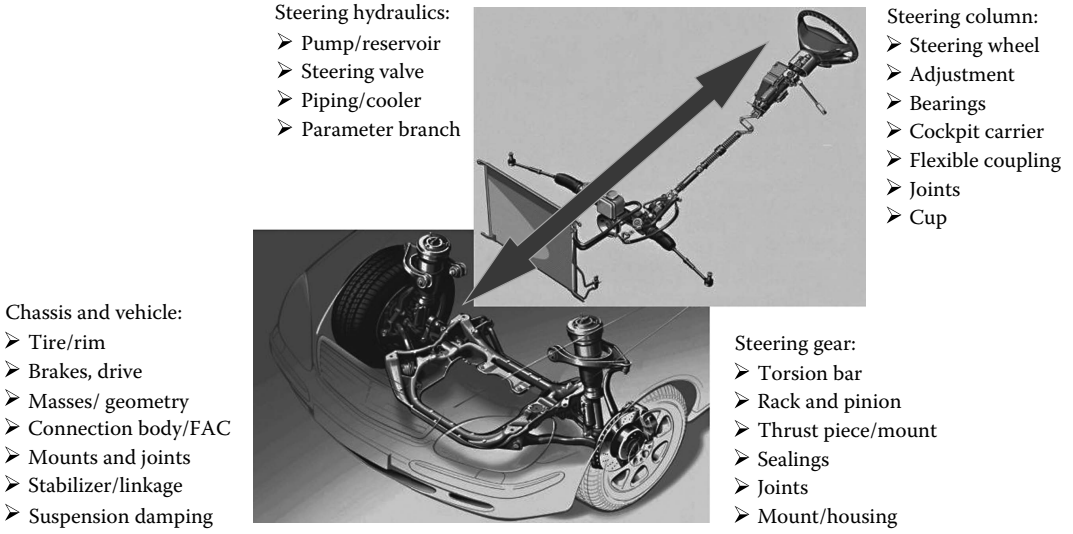


FIGURE 28.27 Hydraulic power steering system of a luxury passenger car (Mercedes S-Class, 2002): major components and parameters influencing the steering behavior.

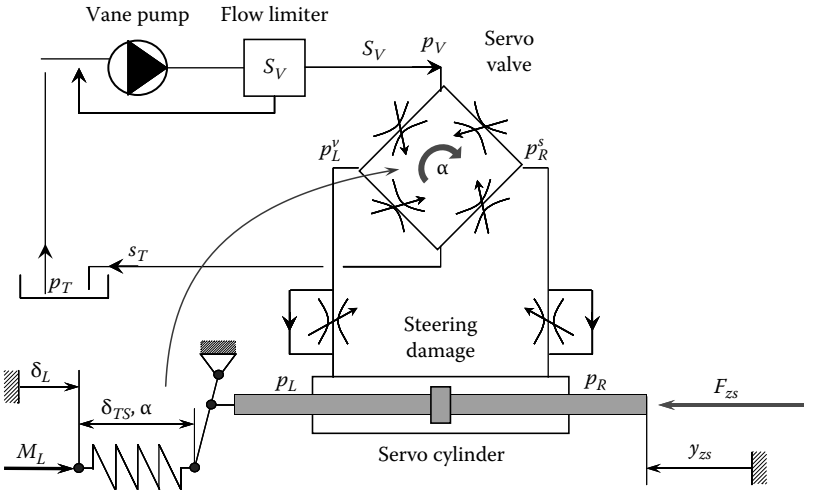


FIGURE 28.28 Basic hydraulics components of a vehicle power steering system (without parameter steering branch/function).

the displacement of the servo valves. These are arranged in a bridge design within the valve block so that valves facing each other diagonally are actuated simultaneously. At a (approximately) constant supply flow rate, the bridge asymmetry results in a pressure difference between cylinder surfaces and therefore to the steering-assisting power. To suppress road and chassis disturbances effectively, unidirectional steering damping valves are installed at the cylinder outlets. These oppose any disturbing excitations directly with matching damping forces. In practice, the damping effect must be restricted, however, in order not to impair quick steering movements by the driver. (For example, during emergency maneuvers, steering wheel rates of up to several hundred degrees per second may be required.)

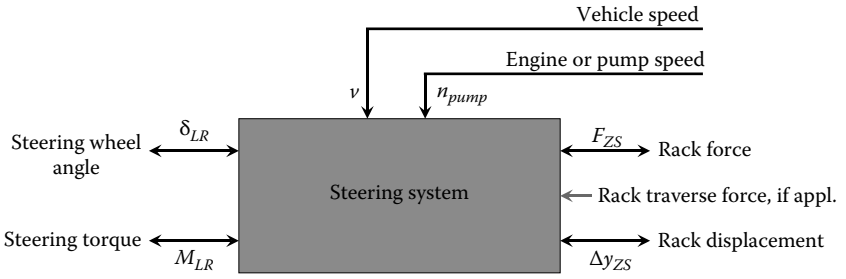


FIGURE 28.29 System limits of a vehicle steering system: four-pole characteristics with primary interaction parameters–steering wheel angle, steering wheel torque, lateral rack displacement, and rack force.

28.6.2 Transfer Characteristics

The steering wheel angle δ_L is probably the primary steering input variable in most driving situations; the steering wheel torque M_L therefore becomes a response or output variable. On transversely inclined roads or under crosswinds, the driver interventions point more toward some type of torque control, and the same applies to sudden interfering road surface excitations [58].

The interaction between the steering system and axle proves similarly complementary: By means of the transverse displacement of the rack y_{zs} relative to the vehicle (or relative to the front axle carrier), the steering imposes the wheel steer angle δ_R on the axle. On closer examination, however, it appears that the steer angles are created from an equilibrium of forces and torques around the chassis steering axles. A further contribution is made by the tie rod forces which are directly related to the transverse rack force F_{zs} of the steering. This also means that no unambiguous input–output behavior can be observed at the interface between the steering and chassis. The steering must therefore be understood and described as a general dynamic four-pole, as illustrated in Figure 28.29.

28.6.3 Steering Models for Vehicle Behavior Analysis

It is expedient to determine the quasi-stationary steering characteristics first for vehicle dynamics and vehicle behavior analyses in order to model the transfer properties for the relevant operating point adequately. Models of this category take into account

- The inertias of the steering wheel and steering column (the inertias of the rack and tie rods are normally added to the degrees of freedom of the axle steering)
- The steering ratio/ratio variations above the steering angle due to inclined joint positions, gearing nonuniformities, or also design gear ratio spreads
- The inner elasticity, damping, and friction of the steering column
- The power assistance, which is normally dependent on the traveling speed

For simple servo valve geometries, the resulting servo forces ΔF_{servo} can be determined directly from the valve resistance values. When friction effects are neglected, a steering torque M_L at a given steering ratio r_L generates a rack force F_{zs} of

$$F_{zs} = \left[\frac{1}{r_L} + \frac{\Delta F_{servo}^{\max}}{M_L^{\max}} \left(\frac{1 - \left(M_L^{\lim} / M_L^{\max} \right)^2}{1 - \left(\min(M_L; M_L^{\lim}) / M_L^{\max} \right)^2} \right)^2 \right] M_L \tag{28.27}$$

whereby ΔF_{servo}^{max} is the maximum servo force, M_L^{max} the related steering wheel torque, and M_L^{lim} the remaining design parameter for the servo characteristic. M_L^{max} is generally hardly smaller than M_L^{lim}. The servo characteristics are therefore very flat around the zero position which reinforces the steering feel during straight-line driving. Vice versa, high steering assistance forces are generated for larger steering torques, for example, during parking. All in all, a very strong nonlinear system characteristic results.

A further problem field in steering analysis and modeling is friction. Friction arises at different locations, among others in the steering column, in the servo valve block; between pinion, rack, and thrust piece; and on the seals and guides of the hydraulic cylinder actuator. Altogether, friction forces of up to 150 N are required, for example, to move the rack with free steering wheel. The various friction points always have design causes, but friction is also used in a targeted way in order to optimize the disturbance transfer behavior of the steering system.

With regard to modeling, several problems result at once: On the one hand, the distributed, sequential friction points in the steering system hardly allow the assignation of the overall effects to the responsible components without splitting up the steering system itself. On the other hand, some of the friction effects are directional, that is, different friction forces are generated for control and disturbance excitations. In addition, some of the friction effects directly depend on the hydraulic pressure—and thereby on the overall system state. (Unfortunately, this similarly applies to electromechanical steering systems.)

28.6.4 Detailed Steering System Models for Axle Dynamics Analysis

Due to these premises, detailed steering models for axle dynamics and vibration analyses are normally sophisticated mechanics–hydraulics compound models with some control extensions, for example, for the representation of parameter steering functions.

The mechanics include, among others, the steering column up to the entry of the steering gear as 1D torsion mechanics or as 3D mechanics model. Steering gears and the steering axle coupling are formulated as 3D mechanics models, in particular, in order to be able to model the effects of the steering mounting in different operational states adequately.

The steering hydraulics must take into account the following components at least:

- Power steering pump
- Flow control and pressure limiter
- Servo valve bridge circuit
- Expansion hoses
- Steering damper valves
- Servo actuator (hydraulic cylinder)

In addition, the modeling of the hydraulic parameter steering components should be considered. This is highly dependent on the manufacturer, however, and cannot therefore be discussed here with general validity. In principle, the parameter or map-controlled steering systems are always identical: A relatively small part of the flow is branched off at pump pressure level p_V; this is accumulated by means of a series of restrictor valves and a relatively slow control valve. The resulting control pressure is used to generate an internal steering torque to reinforce the torsion bar rotation. The required steering torque at low speeds and during parking is thereby reduced (refer to Figure 28.26).

Detailed steering system models are used for the optimization of the steering dynamics and improved adjustment of the disturbance transfer behavior, that is, to avoid steering shimmy, nervous steering, or steering kickback. All adjustment work in this sector is finally targeted at making the tire–road–steering system robust against disturbance excitations while at the same time safeguarding the primary control characteristics (direct, taut steering sensation, freedom from play, etc.).

Looking at the steering kickback phenomenon, that is, excessive steering wheel reactions when driving over bumps while cornering, it becomes clear that a justifiable distribution of the requirements

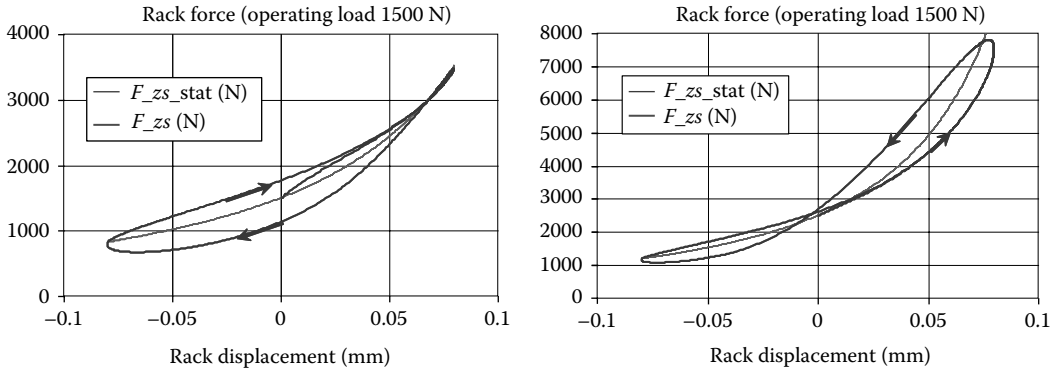


FIGURE 28.30 Disturbance behavior of a power steering system during cornering with high transverse force level, illustrated by a harmonic rack excitation at a frequency of 15 Hz; disturbance force amplitudes of 1.5 kN are absorbed and damped, while disturbance amplitudes of 2.5 kN result in an exciting steering response.

between the steering and chassis must be found or worked out: In this load case, the excitations from the road surface directly generate vertical and transverse axle vibrations resulting in tie rod force fluctuations and therefore steering disturbances. In the event of smaller force excitations or slower force changes, the steering has a damping effect and reduces the exciting energy quickly and efficiently. If the excitation exceeds a certain force or dynamic level, however, the steering system will at first have an exciting effect, as illustrated in Figure 28.30. As, due to the principle involved, this effect can occur in any hydraulic power steering system, the problem is solved conceptually in that the steering, on the one hand, has to cope with a defined dynamic force level by damping and not exciting, and in that the tire–chassis compound on the other hand ensures, in driving situations relevant to bump steering, that this steering limit force and disturbance force level are not exceeded.

28.6.5 Active Steering Systems

In addition to the conventional steering design, different types of active steering systems (torque and superimposed angle steering systems) have been discussed or are in use for some years. A large part of the conventional steering design is normally adopted or retained in these steering systems. The additional functions are realized by means of specific actuators in the steering linkage, as illustrated in Figure 28.31.

Active torque steering systems allow the additional imposition of synthetic steering torques ΔM_L and steering forces ΔF_S, either by appropriately controlling already existing steering actuators or by integrating additional torque actuators (e.g., electric motors) into the steering column. This allows advanced safety functions to be realized, which will, for example, suppress excessive steering movements in critical driving situations. On the other hand, minor chassis weak points such as insufficient steering return can be suppressed. Steering torque actuators also form the basis for a number of comfort functions—from straight-line running assistance via crosswind suppression thru to controlled track guidance based on track recognition sensors.

The force and torque input ensure that the steering intervention assistance should be accepted, given that the interventions are largely not perceivable by the driver and, above all, that the geometrical and dynamic aspects of vehicle perception hardly change as a result of the intervention.

The so-called superimposed steering systems, in contrast, feature additional steering angles. The actuators are arranged in the steering column, either close to the steering wheel or within the power steering between the valve block and gear unit (refer to Figure 28.31).

The primary purpose of superimposed steering systems is to improve vehicle behavior at very low and very high speeds without impairing the middle speed sector, which experience shows to be important. The background to the design problem is the concept-related (stationary) yaw

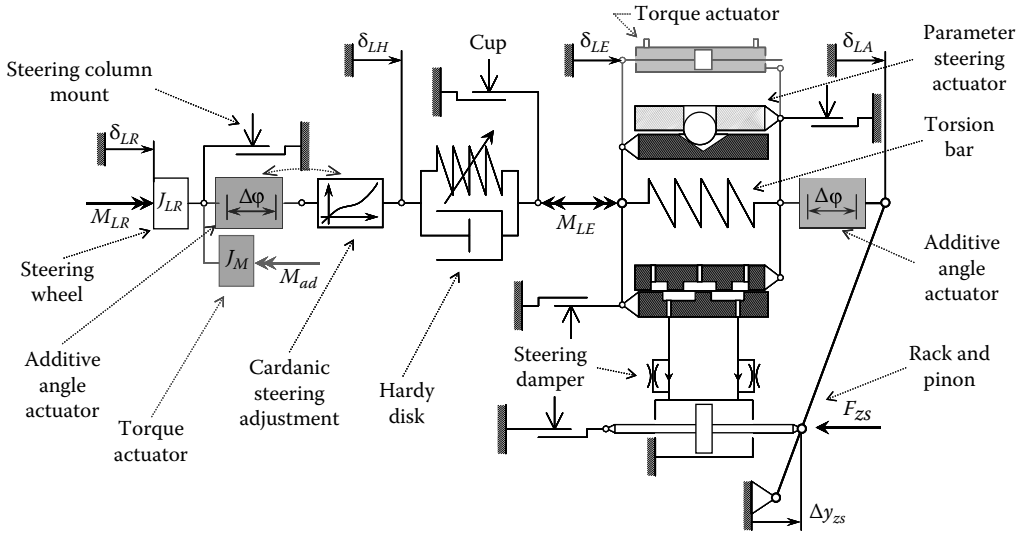


FIGURE 28.31 Structural implementation of superimposed torque (top right) and superimposed angle steering interventions (bottom).

amplification $K_{\dot{\psi}}$ and the transfer of steering angle changes into yaw motions. The yaw velocity $\dot{\psi}$ describes the rotation of the vehicle around its vertical axis:

$$K_{\dot{\psi}} = \left.\frac{\partial\dot{\psi}}{\partial\delta_L}\right|_{stationary} = \frac{v}{l_l + EGv^2}; \quad EG = \frac{m}{l_l}\left(\frac{l_h}{c_v} - \frac{l_v}{c_h}\right) \tag{28.28}$$

To ensure drivability in all speed ranges, the self-steering gradient EG must be set to values between 0.15 and 0.45 Grad/m/s^2. At a given vehicle mass m and geometry (wheelbase l_l, distance from center of gravity – front axle l_v, $l_h = l_l - l_v$), EG is fixed by design by means of the cornering stiffness values c_v and c_h of the front and rear axle. The resulting characteristic for the yaw amplification is illustrated in Figure 28.32.

Conventional vehicles therefore require much larger steering angles at smaller speeds than is desirable. At high speeds, attempts to reduce the sensitivity of the vehicle to rapid steering angle changes to such an extent that completely relaxed and smooth driving is supported are only moderately successful on the other hand. Superimposed steering systems offer the opportunity here, as illustrated in Figure 28.32,

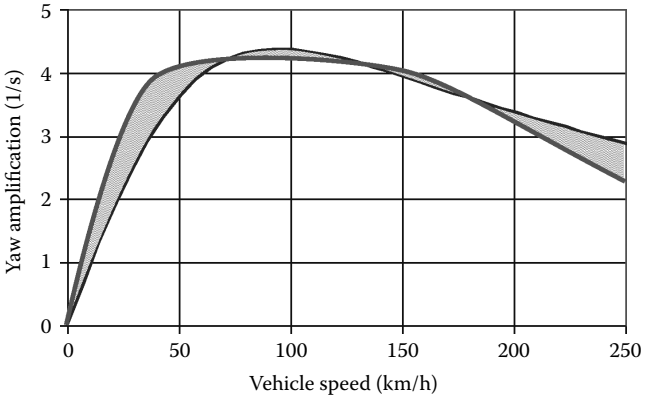


FIGURE 28.32 Yaw amplification for a conventional vehicle (black) and optimized yaw amplification characteristic (grey) on the basis of superimposed steering systems.

to increase the steering ratio significantly at low speeds and at the same time to realize a noticeably more indirect steering system at higher speeds. As long as the superimposed interventions are effected slowly and synchronous to the driver's steering movements, no acceptance problems are to be expected, as was shown in the case of superimposed torque steering systems [61,62]. (It will be hard to convince drivers of rapid superimposed steering interventions, on the other hand, which are not coordinated with the driver; this only seems feasible in the context of safety functions which are rarely activated.)

By combining torque and angle superimposition actuators, the functionality of a steer-by-wire system is recreated. In this case, the steering interventions can be controlled quasi independently of the driver, for example, in order to ensure vehicle stability even under the most adverse conditions or in order to offer the driver different—or also invariant—vehicle characteristics depending on the requirements in particular situations.

REFERENCES

Detailed Modelling, Simulation, and Analysis

1. Milliken, W. F. and Milliken, D. L., *Race Car Vehicle Dynamics*, SAE, Warrendale, PA, 1995.
2. Ammon, D., Apel, A., Mitschke, M., and Schittenhelm, H., Driver behaviour model for longitudinal and directional control in emergency manoeuvres, in: *5th International EAEC-Congress, Conference C Active Safety, Components and Sub-Components*, June 21–23, 1995, Strasbourg, France, SIA-Paper No. 9506C12.
3. Ammon, D., *Modellbildung und Systementwicklung in der Fahrzeugtechnik*, Habil., Universität Karlsruhe (TH), Teubner, Stuttgart, Baden-Württemberg, Germany, 1997.
4. Ammon, D., Gipser, M., Rauh, J., and Wimmer, J., High performance system dynamics simulation of the entire system tire-suspension-steering-vehicle, *Vehicle System Dynamics*, 27(1997), 435–455.
5. Zomotor, Z., Ammon, D., and Meljnikov, D., Modellbasierter fahrversuch, in: VDI-Tagungsbericht *Meß- und Versuchstechnik in der Fahrzeugentwicklung*, Stuttgart, Baden-Württemberg, Germany, 2003, VDI-Ber. Nr. 1755, Düsseldorf, Germany, 2003, pp. 25–36.
6. Rauh, J., Virtual development of ride and handling characteristics for advanced passenger cars, *Vehicle System Dynamics*, 40(1–3), 135–155, 2003.
7. Karnopp, D., *Vehicle Stability*, Marcel Dekker, New York, 2004.
8. Ammon, D., Fahrdynamiksimulationen und Fahrsimulatoren—Potenziale und Grenzen anspruchvoller Entwicklungswerkzeuge, *Motion Simulator Conference*, GZVB, Braunschweig, Lower Saxony, Germany, 2005, pp. 116–135.
9. Valásek, M. (ed.), Computational mechanics in vehicle system dynamics, *Vehicle System Dynamics*, 40, 17–36, Supplement, Taylor & Francis, 2002.
10. Dragon, L., FADYS handling simulations used in a real-time hardware-in-the-loop application with the ESP-controller, *International Symposium on Advanced Vehicle Control AVEC 96*, Vol. 1, Aachen, North Rhine-Westphalia, Germany, 1996, pp. 559–572.
11. Falkner, A. and Reinalter, W., Consistent vehicle model for determining the design envelope, ride comfort and component load, *Vehicle System Dynamics*, 44, 468–478, Supplement, 2006.

Road Modeling

12. Shinozuka, M., Simulation of multivibrate and multidimensional random processes, *Jounal of the Acoustical Society of America*, 49(1971), 357–368.
13. Ammon, D. and Wedig, W., Identification of power spectra—Linear and nonlinear approaches, in: *Proceedings of the ICOSSAR '89 Structural Safety and Reliability*, August 8–11, 1989, San Francisco, CA, ICOSSAR Paper No. T23D-05.
14. Ammon, D., Approximation and generation of Gaussian and Non-Gaussian stationary processes, in: *Proceedings of the Euromech 250 Colloq. Nonlinear Structural Systems under Random Conditions*, June 19–23, 1989, Como, Italy, *Structral Safety*, 8(1990), 153–160.
15. Ammon, D. and Bormann, V., Zur Kohärenz zwischen den Unebenheitsanregungen an der linken und rechten Fahrspur, in: *VDI-Tagung Unebenheiten von Schiene und Straße als Schwingungsursache*, April 23–24, 1991, Braunschweig, Lower Saxony, Germany, VDI Ber. Nr. 877, 1991, 103–118.
16. Ammon, D., *Problems in Road Surface Modelling*, 12th IAVSD-Symposium., Suppl. to *Vehicle System Dynamics*, 20, 28–41, Swets & Zeitlinger, Amsterdam, the Netherlands, 1992.

17. Hofer, R., Lugner, P., and Ammon, D., Safety loss of passenger cars due to road unevennesses, in: *Proceedings of the 6th European Congress on Leightweight and Small Cars—The Answer to Future Needs*, Vol. 1, July 2–4, 1997, Cernobbio, Italy, EAEC-Paper No. 97A2107, pp. 49–59.
18. Gimmler, H., Ammon, D., and Rauh, J., Strassenprofile: Mobile Messung, prozessgerechte Datenaufbereitung und vollständige Bewertung bereiten die Basis für eine effektive Simulation, in: *VDI-Tagungsbericht Reifen-Fahrwerk-Fahrbahn*, Hannover, Lower Saxony, Germany, 2005, VDI-Ber. Nr. 1912, Düsseldorf, North Rhine-Westphalia, Germany, 2005, pp. 335–352.
19. Imine, H., Delanne, Y., and M'Sirdi, N. K., Road profile input estimation in vehicle dynamics simulation, *Vehicle System Dynamics*, 44(4), 285–303, 2006.
20. ISO. 8608, Mechanical vibration—Road surface profiles—Reporting of measured data, Ref.-Nr. ISO 8608:1995(E).
21. Riepl, A., Reinalter, W., and Fruhmann, G., Rough road simulation with tire model RMOD-K and FTire, *Vehicle System Dynamics*, 41, 734–743, Supplement, 2004.

Tire Models

22. Bakker, E., Nyborg, L., and Pacejka, H. B., Tyre modelling for use in vehicle dynamics studies, SAE-Paper No. 870421, 1987.
23. Leister, G., New procedures for tyre characteristics measurement, *Vehicle System Dynamics*, 27, 22–36, Supplement, 1997.
24. Fach, M., Lokale Effekte der Reibung zwischen Pkw-Reifen und Fahrbahn, Diss., VDI-Fortschritts-Ber., Reihe 12, No. 411, 1999.
25. Fischlein, H., Untersuchung des Fahrbahnoberflächeneinflusses auf das Kraftschlussverhalten von Pkw-Reifen, Diss., VDI-Fortschritts-Ber., Reihe 12, No. 414, 1999.
26. Ammon, D., Gnadler, R., Mäckle, G., and Unrau, H.-J., Determination of the friction coefficients of rubber ribs for the exact parameterisation of tyre models, Automobiltechnische Zeitschrift—ATZ, Bd. 106, 2004, Heft 7/8, 694–701.
27. Ammon, D., Vehicle dynamics analysis tasks and related tyre simulation challenges, *Vehicle System Dynamics*, 43, 30–47, Supplement, 2005.
28. Lugner, P. and Plöchl, M. (eds.), Tyre models for vehicle dynamics analysis, *Vehicle System Dynamics*, 43, Supplement, Taylor & Francis, 2005.
29. Oertel, C. and Fandre, A., *Das Reifenmodellsystem RMOD-K, Ein Beitrag zum virtuellen Fahrzeug*, Automobiltechnische Zeitschrift—ATZ, Bd. 103, 2001, Heft 11, 1074–1079.
30. Gipser, M., FTire, A new fast tire model for ride comfort simulations, *International ADAMS User's Conference*, Berlin, Germany, 1999, available from http://www.ftire.com
31. Lugner, P., Pacejka H., and Plöchl, M., Recent advances in tyre models and testing procedures, *Vehicle System Dynamics*, 43(6–7), 413–436, June–July, Taylor & Francis, 2005.
32. Pacejka, H.-B. and Besselink, I. J. M., Magic formula tyre model with transient properties, *2nd International Colloquium on Tyre Models for Vehicle Dynamic Analysis*, Berlin, Germany, February 20–21, 1997, *Vehicle System Dynamics*, 27, 234–249, Supplement, 1997.
33. VERTEC EU Project Final Report: Vehicle, road, tyre and electronic control system interaction—Prediction and validation of handling behaviour, *3rd International Colloquium on Vehicle Tyre Road Interaction, Tyre Technology Conference*, Stuttgart, Baden-Württemberg, Germany, March 8, 2006, www.vertec.hut.fi
34. Besselink, I. J. M., Experiences with the TYDEX standard tyre interface and file format, in: *Tyre Models for Vehicle Dynamics Analysis, Vehicle System Dynamics*, 43, Supplement, Taylor & Francis, 2005, pp. 63–75.
35. Van Oosten, J. J. M., Unrau, H.-J., Riedel, G., and Bakker, E., *TYDEX Workshop: Standardisation of Data Exchange in Tyre Testing and Tyre Modeling*, in: *Tyre Models for Vehicle Dynamic Analysis*, *Vehicle System Dynamics*, 27, Swets & Zeitlinger, Lisse, the Netherlands, 1997, pp. 272–288.
36. Pottinger, M., *Tire Tests for Full Vehicle Road-Load Prediction*, SAE Tire Tests for Road-Load Tire Model Parameter Task Force, 7 March 2006 Tire Expo; J-documents available from www.sae.org: J2704 Tire Normal Force/Deflection and Cross Footprint Dimension Test, J2705 Tire Quasi-Static Envelopment of Triangular/Step Cleats Test, J2710 Modal Testing and Identification of Lower Order Tire Natural Frequencies of Radial Tires, other J-documents in final editing or approval process.
37. Netsch, L., Schick, B., and Wenninger, F., *TÜV—Intelligent tire sensor*, in: *VDI-Tagungsbericht Reifen-Fahrwerk-Fahrbahn*, Hannover, Lower Saxony, Germany, 2005, VDI-Ber. Nr. 1912, Düsseldorf, North Rhine-Westphalia, Germany, 2005, pp. 153–167.
38. Narasimha Rao, K. V., Krishne Kumar, R., Mukhopadhyay, R., and Misra, V. K., A study of relationship between different Magic Formula coefficients and tyre design attributes through finite element analysis, *Vehicle System Dynamics*, 44(1), 33–63, 2006.

39. Gäfvert, M. and Svendenius, J., A novel semi-empirical tyre model for combined slips, *Vehicle System Dynamics*, 43(5), 351–384, 2005.
40. Rauh, J. and Mössner-Beigel, M., Tire simulation challenges, Paper #104 to be published on the *20th IAVSD-Symposium*, Berkeley, CA, 2007.

Axles and Chassis Components

41. Bastow, D., Howard, G., and Whitehead, J. P., *Car Suspension and Handling*, 4th edn., Pentech Press, London, U.K., 2004.
42. Sjöberg, M., On dynamic properties of rubber isolators, Ph.D. Thesis, KTH, Stockholm, Sweden, 2002.
43. Matschinsky, W., *Road Vehicle Suspensions*, Wiley Publishers, New York, 1997.
44. Miege, A. J. P. and Cebon, D., Optimal roll control of an articulated vehicle: theory and model validation, *Vehicle System Dynamics*, 43(12), 867–893, 2005.

Suspension Systems

45. Ammon, D., New perspectives in vehicle dynamics—The mechatronic suspension system of the F400 Carving, in: *VDA-Kongreß Sicherheit durch Elektronik*, März 20.–21, 2002, VDA, Frankfurt, Germany, pp. 345–354.
46. Ammon, D., Gimmler, H., Hilf, K.-D., Rauh, J., Scheible, G., and Stiess, P., *Fahrzeugschwingungen—Von der Fahrbahnanregung bis zum Komfortempfinden*, VDI-Tagung *Humanschwingungen*, Darmstadt, Hesse, Germany, 2004, VDI-Ber. Nr. 1821, Düsseldorf, North Rhine-Westphalia, Germany, 2004, pp. 139–158.
47. Schiehlen, W. O. and Schirle, T., Modeling and simulation of hydraulic components for passenger cars, *Vehicle System Dynamics*, 44, 581–589, Supplement, 2006.
48. Yung, V. Y. B. and Cole, D. J., Modelling high frequency force behaviour of hydraulic automotive dampers, *Vehicle System Dynamics*, 44(1), 1–31, 2006.
49. Dixit, R. K. and Buckner, G. D., Sliding mode observation and control for semiactive vehicle suspensions, *Vehicle System Dynamics*, 43(2), 83–105, 2005.
50. Smith, M. C. and Wang, F-C., Performance benefits in passive vehicle suspension employing inerters, *Vehicle System Dynamics*, 42(4), 235–257, 2004.
51. Strandemar, K. and Thorwald, B., Ride diagram: A tool for analysis of vehicle suspension settings, *Vehicle System Dynamics*, 44, 913–920, Supplement, 2006.
52. Bosch, R., *Automotive Handbook*, 6th edn., Robert Bosch GmbH, Plochingen, London, U.K., 2004.
53. Schöneburg, R. and Breitling, T., Enhancement of active & passive safety by future pre-safe systems, *19th ESV Conference*, Washington, DC, 2005, Paper # 05-0080.
54. Liebemann, E. K., Meder, K., Schuh, J., and Nenninger, G., Safety and performance enhancement: the bosch electronic stability control (ESP), *19th ESV Conference*, Washington, DC, 2005, Paper # 05-0471.

Steering Systems

55. Groll, M. V., Müller, S., Meister, T., and Tracht, R., Disturbance compensation with a torque controllable steering system, *Vehicle System Dynamics*, 44(4), 327–338, 2006.
56. Tseng, H. E., Asgar, J., Hrovat, D., van der Jagt, P., Cherry, A., and Neads, S., Evasive manoeuvres with a steering robot, *Vehicle System Dynamics*, 43(3), 199–216, 2005.
57. Yuhara, N. and Tajima, J., Advanced steering system adaptable to lateral control task and driver's intention, *Vehicle System Dynamics*, 36(2–3), 119–158, 2001.
58. Ammon, D., Börner, M., and Rauh, J., Simulation of the perceptible feed-forward and feed-back properties of hydraulic power steering systems on the vehicle's handling behaviour using simple physical models, *Vehicle System Dynamics*, 44, 158–170, Supplement, 2006.
59. Bröcker, M., New control algorithms for steering feel improvements of an electric powered steering system with belt drive, *Vehicle System Dynamics*, 44, 759–769, Supplement, 2006.
60. Chai, Y. W., Abe, Y., Kano, Y., and Abe, M., A study on adaptation of SBW parameters to individual driver's steer characteristics for improved driver-vehicle system performance, *Vehicle System Dynamics*, 44, 874–882, Supplement, 2006.
61. Neukum, A. and Krüger, H. P., Fahrerreaktionen bei Lenksystemstörungen—Untersuchungsmethoden und Bewertungskriterien, VDI-Bericht-Nr. 1791, 2003, pp. 297–318.
62. Reinelt, W. and Lundquist, C., Controllability of active steering system hazards: From standards to driving tests, in: *Proceedings of the SAE World Congress*, Detroit, MI, April 2006, SAE paper 2006-01-0600.
63. Ammon, D. and Schiehlen, W., Advanced road vehicles: Control technologies, driver assistance, in: *Dynamical Analysis of Vehicle Systems*, CISM courses and lectures, Vol. 497, Springer, 2007, pp. 283–304.

29 Ride Comfort and Road Holding

*Karl Popp**

CONTENTS

29.1 Comfort and Safety Criteria 1059
29.1.1 Driving Stability 1060
29.1.2 Ride Comfort 1061
29.1.2.1 Deterministic Excitation 1061
29.1.2.2 Stochastic Excitation 1064
29.1.3 Ride Safety 1066
29.2 Modeling of Random Vehicle Excitation 1067
29.2.1 Mathematical Description of Random Processes 1068
29.2.2 Models of Road Roughness Profiles 1077
29.2.3 Models of Vehicle Excitation 1079
29.3 Computation of the Random Vehicle Response 1082
29.3.1 Numerical Simulation 1083
29.3.2 Spectral Density Analysis 1085
29.3.3 Covariance Analysis 1087
References 1091

29.1 COMFORT AND SAFETY CRITERIA

The assessment of dynamical vehicle properties is related to the three essential motions:

- Longitudinal motion (driving and braking)
- Lateral motion (guidance and steering)
- Vertical motion (suspension and damping)

These three motions of a vehicle will be considered decoupled for the sake of simplicity so that the assessment criteria can be formulated separately for each motion. Of course, coupled criteria can be developed if necessary. But then, one has to compete with the subjective impression during driving tests. Thus, the question of a mathematical formulation of assessment criteria is difficult and can often be answered empirically only (see Chapter 37). Sometimes an interdisciplinary approach is helpful taking into account ergonomical knowledge.

The criteria to assess the longitudinal motion are known as *vehicle performance*. It comprises maximum speed, resistance of motion, and vehicle acceleration. These criteria are self-explaining to some extent (see also Chapter 27). However, it should be mentioned that the exact measurement of the vehicle performance in driving tests is a nontrivial problem (see Chapter 7).

The essential criterion for the lateral motion is the *driving stability*. The vehicle should safely follow a trajectory prescribed by the steering systems. Thus, a stability problem is given, which can be solved by applying the well-developed methods of system dynamics (see Chapter 5).

* Author Karl Popp passed away unexpectedly on 24 April 2005. This Chapter was respectfully revised by Werner Schiehlen.

For the vertical motion, the criteria *ride comfort and safety* are essential. The wheel suspension should absorb the disturbances coming from the road irregularities so that a comfortable ride becomes possible. At the same time, the weight of the vehicle should be transferred to the road without wheel vibrations, so that large lateral/longitudinal forces in the contact area remain in order to guarantee safety.

Besides the vehicle motions, the static and dynamic forces acting on the vehicle parts are important criteria, because the *lifetime* of the vehicle components depends strongly on the loads (see Chapters 9 and 10). These loads can be calculated in advance by model-based numerical simulations. In the following sections, the aforementioned criteria, driving stability, ride comfort, and safety, will be considered in detail. Scalar measures are introduced to simplify the assessment.

29.1.1 Driving Stability

The lateral motion generally can be described by homogeneous linear or nonlinear state equations,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) \tag{29.1}$$

or

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{a}(\boldsymbol{x}, t)\,, \boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{29.2}$$

Thus, the driving stability can be defined as stability of motion in the sense of Lyapunov, which means:

The equilibrium state $\boldsymbol{x}(t) = \boldsymbol{0}$ of the dynamical system (29.2) with $\boldsymbol{a}(\boldsymbol{0}, t) = \boldsymbol{0}$ is called stable in the sense of Lyapunov, if for each t_0 and all $\varepsilon > 0$, there exists a positive number $\delta = \delta(\varepsilon, t_0) > 0$, such that for an arbitrary initial state $\boldsymbol{x}_0$ with

$$\| \boldsymbol{x}_0 \| < \delta \tag{29.3}$$

it follows a disturbed motion with

$$\| \boldsymbol{x}(t) \| < \varepsilon, t \geq t_0. \tag{29.4}$$

Here, $\|\boldsymbol{x}\|$ means any norm of the vector $\boldsymbol{x}$. Often, the Euclidian norm $\| \boldsymbol{x} \| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$ is chosen. Furthermore, if

$$\lim_{t \to \infty} \boldsymbol{x}(t) = 0, \tag{29.5}$$

holds, then the system is called asymptotically stable. On the other hand, the equilibrium state $x(t) = 0$ is called unstable, if it is not stable.

Example 29.1

Stability behavior of a second-order system.

For a system with one degree-of-freedom or two state variables x_1, x_2 that are combined in the state vector $\boldsymbol{x}$, $\boldsymbol{x} = [x_1\ x_2]^{\mathrm{T}}$, the trajectories shall be visualized in the phase plane (see Chapter 5) in case of an asymptotically stable, a stable, and an unstable equilibrium position, respectively.

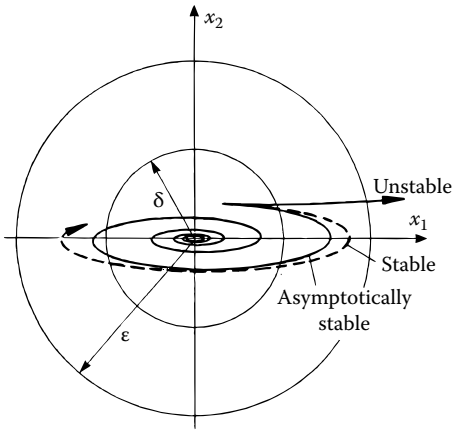


FIGURE 29.1 Solutions of a second-order system plotted in the phase plane.

Solution:

For a second-order system, the solution can be visualized completely in the phase plane (see Chapter 5) (see Figure 29.1). At first, the neighborhood ε of the equilibrium state, that is the origin of the coordinates, will be chosen. Then the neighborhood δ will be determined in such a way that the stable and asymptotically stable trajectories remain completely within the ε–neighborhood. This means that the δ–neighborhood should possibly be very small. Here, the Euclidian norm

$$|| \boldsymbol{x} || = \sqrt{x_1^2 + x_2^2} \tag{29.6}$$

has been applied, resulting in circles around the origin in the phase plane.

The definition of stability in the sense of Lyapunov holds also for linear systems (29.1). This definition, however, allows the assessment of stability only if the solution of the system is known. Thus, stability criteria that do not require the explicit solution of the system are of special interest. Such stability criteria are especially well developed for linear systems [1].

29.1.2 Ride Comfort

Ride comfort (see Chapter 36) is based on the subjective human perception. Numerous experimental investigations in ergonomics have shown that the *human perception* of vibrations depends on the acceleration,

$$K = K(a), \tag{29.7}$$

where

K is a nondimensional *perception measure*

a is the absolute value of the acceleration in horizontal or vertical direction, respectively, see Figure 29.2

Furthermore, the position of the human body (sitting, standing, or lying) is of importance. In ergonomics, the vibration tests have been performed using deterministic, especially harmonic, excitations. In vehicles, however, the human body usually is exposed to random vibrations. Therefore, in the following, the different types of excitations will be investigated separately.

29.1.2.1 Deterministic Excitation

For a rough qualitative assessment of ride comfort during a deterministic excitation,

$$K \sim a_{\max} \tag{29.8}$$

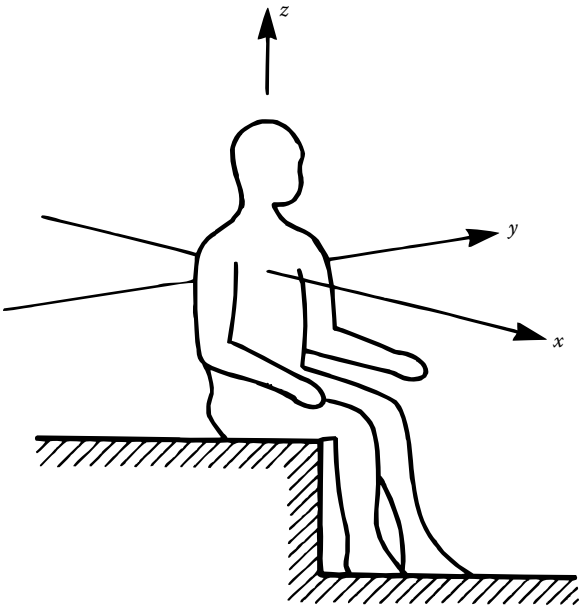


FIGURE 29.2 Directions of vibrations acting on a seated person.

is assumed, that is, the maximum acceleration value prevails upon the perception measure. As a thumb rule, $a_{\max} \leq 0.5$ m/s^2 results in good ride comfort. More refined relations between K and a are given in the international or national standards that are based on extensive ergonomical investigations (see ISO-Standard 2631 [2] or VDI-Standard 2057 [3]).

For harmonic excitations, the aforementioned standards give a precise relation that suits well for vehicle dynamics. Starting from the harmonic acceleration

$$a(t) = A\sin\omega t, \quad \omega = 2\pi f, \tag{29.9}$$

with frequency f [Hz], the root mean square (rms) value a_{rms} of the acceleration can be determined,

$$a_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T a^2(t)\,\mathrm{d}t} = \frac{A}{\sqrt{2}}\ [\text{m/s}^2], \tag{29.10}$$

where the unit m/s^2 has to be used. Then the perception measure K reads

$$\begin{aligned} K &= 10\,a_{\text{rms}}\sqrt{f}, && 1 \leq f \leq 4 \\ K &= 20\,a_{\text{rms}}, && 4 \leq f \leq 8, \\ K &= 160\ a_{\text{rms}} f^{-1}, && 8 \leq f \leq 80, \end{aligned} \tag{29.11}$$

for vertical excitation, and

$$\begin{aligned} K &= 28\,a_{\text{rms}}, && 1 \leq f \leq 2, \\ K &= 56\,a_{\text{rms}} f^{-1}, && 2 \leq f \leq 80. \end{aligned} \tag{29.12}$$

for horizontal excitation.

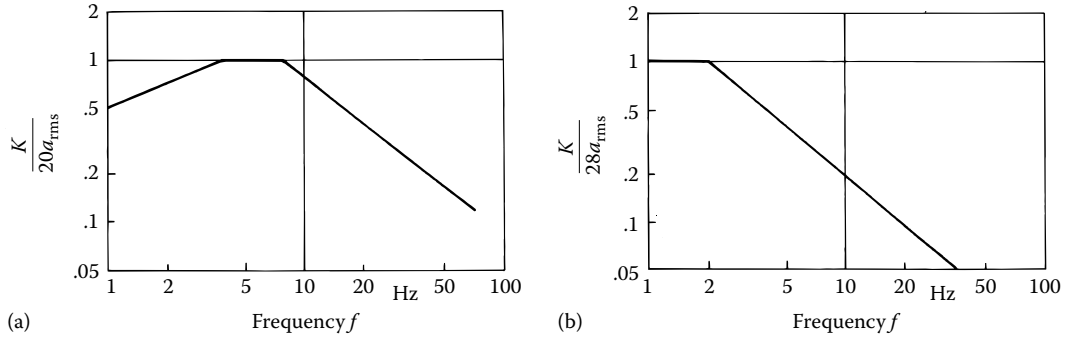


FIGURE 29.3 Frequency response functions of human perception for (a) vertical and (b) horizontal direction.

Equations 29.11 and 29.12 can be interpreted as experimentally determined frequency response functions (FRFs) of the human perception; they are depicted in Figure 29.3. From Figure 29.3a, one can observe that a vertical excitation between 4 and 8 Hz is perceived as very unpleasant, because resonance of the human stomach occurs in this range. The perception K is a nondimensional measure characterizing the subjective impression of test persons related to the criteria comfort, efficiency, and health respectively. Here, the exposure time plays an essential role. The standards ISO 2631 and VDI 2057 describe the relation between exposure time and the aforementioned limits (see Figure 29.4). The perception measure for road vehicles lies in the range

$$2 < K < 10. \tag{29.13}$$

Subclasses of the perception can be defined and the corresponding vibration levels can be described:

$$\begin{aligned} &C_1/C_2, \quad \text{noticeable,} \\ &D_1/D_2, \quad \text{strong noticeable, and} \\ &E_1/E_4, \quad \text{very strong noticeable.} \end{aligned} \tag{29.14}$$

Obviously, the description of perception shows some uncertainties, due to the subjective human nature.

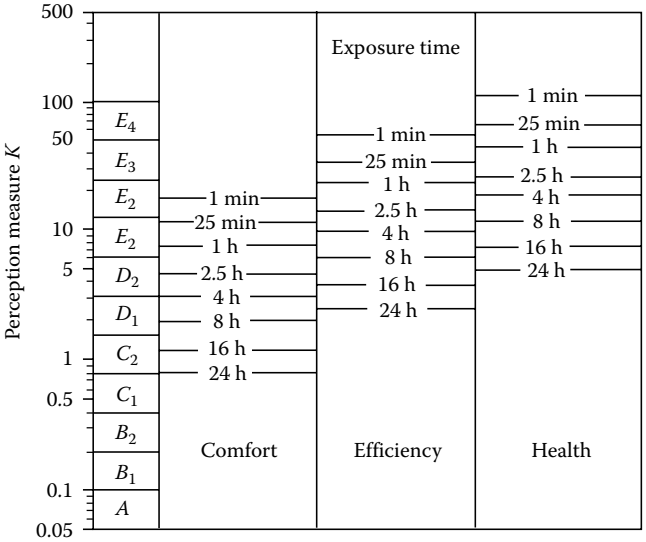


FIGURE 29.4 Human perception and exposure time of vibrations.

29.1.2.2 Stochastic Excitation

Experimental results for a stochastic excitation of test persons are very limited (see e.g., Chapter 22). However, they confirm the interpretation of the subjective human perception as a response to stochastic excitation also, in the sense of a mechanical system, and the possibility of describing it using the methods of linear system theory.

The rms value of a Gaussian process $a(t)$ is equivalent to its standard deviation σ_a,

$$a_{\mathrm{rms}} = \sigma_a = \sqrt{\int_0^\infty \Phi_a(\omega)\,\mathrm{d}\,\omega}, \tag{29.15}$$

where $\Phi_a(\omega)$ is the single-sided power spectral density (PSD) of the process under consideration. The standard deviation is a scalar parameter characterizing the stochastic process globally; however, it is frequency independent. Thus, the weighting of frequencies and the calculation of the rms value have to be interchanged, which is allowed for linear stochastic systems. This results in the variance $\sigma_{\bar{a}}^2$ of the weighted process $\bar{a}(t)$,

$$\sigma_{\bar{a}}^2 = \int_0^\infty \alpha^2 \,|\, F(\omega) \,|^2\, \Phi_a(\omega)\,\mathrm{d}\,\omega. \tag{29.16}$$

Here, α is a dimensional factor, $F(\omega)$ is the FRF of a weighting filter, and $\Phi_a(\omega)$ is the single-sided PSD of the acceleration process $a(t)$. The still unknown quantities of the frequency weighting in Equation 29.16 can be found by comparison with the results for deterministic excitation. For vertical excitation, the dimensional factor α reads

$$\alpha = 20\ \mathrm{s}^2/\mathrm{m}, \tag{29.17}$$

and the FRF is given by

$$\left|F(\omega)\right| = \frac{1}{20} K(a_{\mathrm{rms}}, f)\,,\ f = \frac{\omega}{2\pi}, \tag{29.18}$$

where the perception K according to Equation 29.11 has to be introduced.

The FRF (29.18) can be approximated well by a linear weighting filter or shape filter. This yields in frequency domain

$$\begin{aligned} F(\omega) &= \bar{\boldsymbol{h}}^{\mathrm{T}} (i\omega\, \boldsymbol{E} - \bar{\boldsymbol{F}})^{-1} \bar{\boldsymbol{g}} \\ &= \frac{b_0 + b_1(i\omega) + \cdots + b_r(i\omega)^r}{a_0 + a_1(i\omega) + \cdots + a_{s-1}(i\omega)^{s-1} + (i\omega)^s} \end{aligned} \tag{29.19}$$

Here, $\bar{\boldsymbol{h}}$ and $\bar{\boldsymbol{g}}$ are $s \times 1$-vectors and $\bar{\boldsymbol{F}}$ is a $s \times s$ matrix characterizing the shape filter. Furthermore, a_i, $i = 0(1)s-1$ and b_j, $j = 0(1)r < s$ are coefficients of the shape filter FRF. The following relation holds,

$$\bar{\boldsymbol{F}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{s-1} \end{bmatrix}, \quad \bar{\boldsymbol{g}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \bar{\boldsymbol{h}}^{\mathrm{T}} = [b_0\, b_1 \ldots b_r\, 0 \ldots 0], \tag{29.20}$$

that is, the system matrix $\bar{\boldsymbol{F}}$ can be chosen as Frobenius matrix.

In the time domain, the shape filter can be described equivalently by a differential equation,

$$\begin{aligned} \bar{a}(t) &= \alpha\, \bar{\boldsymbol{h}}^{\mathrm{T}} \bar{\boldsymbol{v}}(t), \quad \text{and} \\ \dot{\bar{\boldsymbol{v}}}(t) &= \bar{\boldsymbol{F}}\, \bar{\boldsymbol{v}}(t) + \boldsymbol{g} a(t). \end{aligned} \tag{29.21}$$

Here, $\bar{\boldsymbol{v}}$ is the $s \times 1$-state vector of the shape filter that is excited by the acceleration $a(t)$. The scalar product of the $s \times 1$-vectors $\bar{\boldsymbol{h}}$ and $\bar{\boldsymbol{v}}(t)$ results in the frequency weighted scalar acceleration $\bar{a}(t)$. The coefficients of the shape filter have to be determined in such a way that the given FRF (29.18) is approximated sufficiently well. This can always be achieved by choosing the order s of the filter large enough.

Example 29.2

Shape filter for the human perception.

A second-order shape filter shall be determined so that the FRF (29.11) for the human perception of vibration is approximated well.

Solution:

The shape filter of second order, $s = 2$, has the following form according to (29.20) and (29.21):

$$\bar{a} = \alpha [b_0 \; b_1] \begin{bmatrix} \bar{v}_1 \\ \bar{v}_2 \end{bmatrix}, \tag{29.22}$$

$$\begin{bmatrix} \dot{\bar{v}}_1 \\ \dot{\bar{v}}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} \bar{v}_1 \\ \bar{v}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} a(t). \tag{29.23}$$

Choosing the numbers

$$\begin{aligned} a_0 &= 1200\ \mathrm{s}^{-2}, \quad b_0 = 500\,\mathrm{s}^{-2}, \\ a_1 &= 50\ \mathrm{s}^{-1}, \quad b_1 = 50\,\mathrm{s}^{-1}, \\ \alpha &= 20\ \mathrm{s}^2\ \mathrm{m}^{-1} \end{aligned}$$

in case of vertical excitation results in the FRF depicted in Figure 29.5. It can be seen that already a second-order shape filter approximates the standard ISO 2631 surprisingly well. The deviations remain completely within the tolerances allowed by the standard.

Similarly, a second-order shape filter regarding the frequency weighting (29.12) for the horizontal excitation can be determined using the numbers

$$\begin{aligned} a_0 &= 75\ \mathrm{s}^{-2}, \quad b_0 = 31.25\,\mathrm{s}^{-2}, \\ a_1 &= 12.5\,\mathrm{s}^{-1}, \quad b_1 = 12.5\,\mathrm{s}^{-1}, \\ \alpha &= 28\ \mathrm{s}^2\ \mathrm{m}^{-1}. \end{aligned}$$

Thus, the computation of the perception quantity K results simply in

$$K = \sigma_{\bar{a}}, \tag{29.24}$$

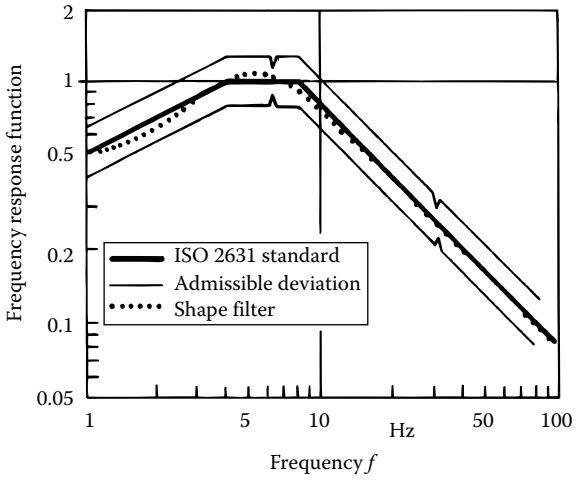


FIGURE 29.5 Shape filter FRF of human perception in the vertical direction.

which includes the frequency weighting automatically. Using the *K*-value from Equation 29.26 again, the allowed exposure time for the criteria comfort, efficiency, and health follows from Figure 29.4.

29.1.3 Ride Safety

Ride safety is related to the longitudinal braking forces and the lateral guidance forces of a vehicle. The tangential contact forces between tire and road are influenced essentially by the normal forces in the contact area. Hence, the remaining minimum wheel load is a criterion for ride safety.

The total wheel load f_{Wheel} can be divided into a static load f_{Stat} due to the vehicle weight and a dynamic load f_{Dyn} generated by vehicle vibrations,

$$f_{\text{Wheel}}(t) = f_{\text{Stat}} + f_{\text{Dyn}}(t), \tag{29.25}$$

as visualized in Figure 29.6. The remaining minimum total wheel load reads

$$f_{\text{Wheel, min}} = f_{\text{Stat}} - f_{\text{Dyn, max}}. \tag{29.26}$$

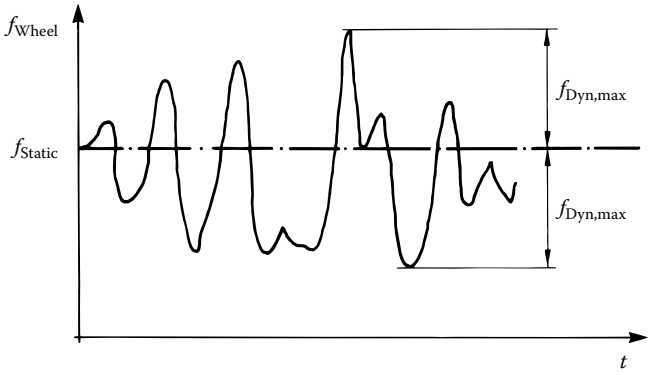


FIGURE 29.6 Static and dynamic wheel loads.

This defines the *safety margin R*,

$$R = \frac{f_{\text{Stat}} - f_{\text{Dyn, max}}}{f_{\text{Stat}}}, \tag{29.27}$$

a scalar quantity varying in the range of $0 \leq R \leq 1$. The safety margin for a vehicle in standstill is $R = 1$; it can drop down to $R = 0$ for a fast-moving vehicle on a rough road.

For roads with a randomly disturbed surface, the dynamic wheel load variations can be determined from numerical simulations. For an assessment of the ride safety, the standard deviation σ_f of the vertical wheel load variations can be used.

This yields

$$R = \frac{f_{\text{Stat}} - \sigma_f}{f_{\text{Stat}}}, \tag{29.28}$$

where R has to be considered as a statistical quantity.

The computation of dynamical wheel load variations, that is, f_{Dyn} or σ_f, is based on the state and input quantities of the global vehicle guideway system. Depending on the tire model, the relative displacements and velocities also have to be known. Hence, the following general expressions can be given,

$$f_{\text{Dyn}} = \boldsymbol{b}^{\text{T}} \boldsymbol{x} \tag{29.29}$$

or

$$\sigma_f^2 = \boldsymbol{b}^{\text{T}} \boldsymbol{P}_x \boldsymbol{b}, \quad \boldsymbol{P}_x = E\{\boldsymbol{x}\,\boldsymbol{x}^{\text{T}}\} \tag{29.30}$$

where $\boldsymbol{b}$ denotes a $n \times 1$ vector of coefficients; $\boldsymbol{x}(t)$, the $n \times 1$-state vector; and $\boldsymbol{P}_x$ the corresponding $n \times n$-covariance matrix.

29.2 MODELING OF RANDOM VEHICLE EXCITATION

Vehicles on rigid guideways are subject to perturbations due to various reasons which have to be added to the vehicle models [4]. Generally, individual obstacles and the guideway unevenness can be distinguished. The obstacles comprise steps (rail joints curbstone edges) and cleats (potholes and bumps) that can be described mathematically by step and impulse functions, primarily resulting in free vibrations of the vehicle. The guideway unevenness causes forced vibrations of the vehicle. The unevenness is a permanent and, therefore, very important source of excitation. In the past, it has been assumed sometimes to be sinusoidal and, thus, it was treated deterministically as well. Nowadays, the unevenness is regarded as random [5], and stochastic methods are applied. In principle, the guideway unevenness could be measured locally with sufficient accuracy and reproducibility, and be treated deterministically. Nevertheless, the main focus in vehicle dynamics is the consideration of a variety of local unevenness representing globally roads or tracks. For this purpose, stochastic processes provide suitable mathematical methods. In the following, the guideway irregularities will at first be described as a *roughness profile* depending on a spacial coordinate. Then, with regard to vehicle speed, it will be transformed into time domain, resulting in a *random excitation process* acting on the vehicle and generating random vibration. The resulting system response is important to assess, for example, ride comfort and ride safety.

The investigation of road roughness profiles has been an ongoing process for decades [6–9]. Numerous measurements have shown that a mathematical description can be given in the form of

- Stationary, normal distributed, and ergodic random processes.

Prior to the description of roughness profiles and the corresponding excitation processes, a general introduction to random processes shall be given in order to explain the aforementioned properties (see also [10–12]).

29.2.1 Mathematical Description of Random Processes

Let us consider a special type of road, for example, the network of highways in a country, and let us assume that the roughness has been recorded by a large number of sectional profile measurements $\zeta^{(r)}(x)$, $r = 1, 2, \ldots$, that are put together in a diagram one above the other (see Figure 29.7a). The independent variable x denotes the distance from an arbitrary chosen starting point in each section. The single roughness profiles are different from each other, $\zeta^{(r)}(x) \neq \zeta^{(s)}(x)$ for $r \neq s$. The complete set of measurement records, however, builds a random process $\zeta(x)$ that represents the highway roughness. Each single record $\zeta^{(r)}(x)$ is called a *realization or sample function* of the process. The ensemble of roughness values $\zeta_j = \zeta(x_j)$ for a definite distance $x = x_j$, $j = 1, 2, \ldots$, builds a random *variable*. It can statistically be evaluated resulting in the *probability density function* $p(\zeta)$ (Figure 29.7b). From $p(\zeta)$ follows the *probability* Pr that the random variable ζ_j lies within a certain interval $[a, b]$,

$$\Pr(a \le \zeta_j \le b) = \int_a^b p(\zeta_j)\,\mathrm{d}\zeta_j, 0 \le \Pr \le 1. \tag{29.31}$$

If the interval grows to infinite, then the roughness value ζ_j lies in it for sure, that is, the probability is Pr = 1,

$$\int_{-\infty}^{\infty} p(\zeta_j)\,\mathrm{d}\zeta_j = 1. \tag{29.32}$$

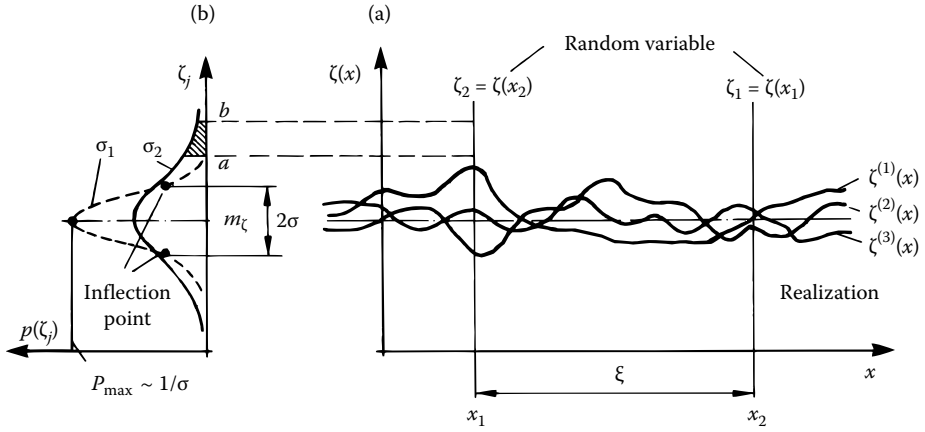


FIGURE 29.7 Roughness profiles as a random process, (a) realization $\zeta(r)(x)$ and random variable ζ_j, (b) probability density function $p(\zeta_j)$.

The *moments* of the probability density function are called mathematical expectation $E\{\cdot\}$; they characterize the random variable. The most important ones are the first-order moment called mean $m_\zeta(x_j)$ or *expected value* and the second-order moment called *mean square value* $m_{\zeta^2}(x_j)$.

For $j = 1$, it follows

$$m_\zeta(x_1) = E\left\{\zeta(x_1)\right\} \equiv \int_{-\infty}^{\infty} \zeta_1 p(\zeta_1)\,\mathrm{d}\zeta_1, \tag{29.33}$$

$$m_{\zeta^2}(x_1) = E\left\{\zeta^2(x_1)\right\} \equiv \int_{-\infty}^{\infty} \zeta_1^2 p(\zeta_1)\,\mathrm{d}\zeta_1. \tag{29.34}$$

The square root of (29.34) is called *root mean square* value or *rms* value. An important statistical parameter is the variance $P_\zeta(x_1) \equiv \sigma_\zeta^2(x_1)$ or the square of the standard deviation $\sigma_\zeta(x_1)$. It is the ensemble average of the square of the deviation from the mean, the so-called central second moment,

$$P_\zeta(x_1) \equiv \sigma_\zeta^2(x_1) = E\left\{\left[\zeta(x_1) - m_\zeta(x_1)\right]^2\right\} = E\left\{\zeta^2(x_1)\right\} - m_\zeta^2(x_1). \tag{29.35}$$

The mathematical exception $E\{\cdot\}$ is a linear operation, where $E\{c\} = c$ for $c = \text{const}$. This property has been used in Equation 29.35. First- and second-order moments are well known from the determination of the center of mass and the moment of inertia. For readers not familiar with the concept of mean and variance, the *mechanical mass analogy* may be helpful. A rod or wire is imagined with infinite length that has a variable mass density ρ (mass per unit length) identical to the probability density $p(x)$ in the probability distribution. According to Equation 29.32, the total mass of the rod is normalized to unity. The mean value (29.33) corresponds to the coordinate of the center of mass and the mean square value (29.34) can be interpreted as moment of inertia related to the origin. The variances (29.35), however, correspond to the moment of inertia related to the center of mass. Thus, the relation $\sigma_\zeta^2 = E\{\zeta^2\} - m_\zeta^2$ in Equation 29.35 can be interpreted as the parallel axes theorem. The first two moments contain information about the distribution of the corresponding quantities, that is, probability or mass, respectively.

Let us consider two random variables $\zeta_1 = \zeta(x_1)$ and $\zeta_2 = \zeta(x_2)$ of the random process ζ. From the statistical relationship between these variables, insight is gained into the random process itself. This is achieved by the (auto) *correlation* $R_\zeta(x_1, x_2)$ or the (auto) *covariance* $P_\zeta(x_1, x_2)$, where, in the latter, the central second moments are used,

$$R_\zeta(x_1, x_2) = E\left\{\left[\zeta(x_1)\zeta(x_2)\right]\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \zeta_1\zeta_2 p(\zeta_1, \zeta_2)\,\mathrm{d}\zeta_1\,\mathrm{d}\zeta_2, \tag{29.36}$$

$$\begin{aligned} P(x_1, x_2) &= E\left\{\left[\zeta(x_1) - m_\zeta(x_1)\right]\left[\zeta(x_2) - m_\zeta(x_2)\right]\right\} \\ &= E\left\{\zeta(x_1)\zeta(x_2)\right\} - m_\zeta(x_1)m_\zeta(x_2). \end{aligned} \tag{29.37}$$

Here, $p(\zeta_1, \zeta_2)$ denotes the second-order probability density function. The prefix "auto" refers to the fact that both random variables belong to the same random process. Otherwise, that is, if the two random variables belong to different random processes, the prefix "cross" will be used. In the limit

case $x_1 = x_2$ of equal correlation positions, the correlation (29.36) becomes equal to the mean square value (29.34) and the covariance (29.37) becomes identical to the variance (29.35).

After these prerequisites, the properties stationarity, normal or Gaussian distribution, and ergodicity shall be explained and described mathematically.

A random process is called *stationary* (or homogeneous) if its statistical properties are invariant under a shift of the origin of the x axis. Thus, all first-order probability density functions are equal, $p(\zeta_1) = p(\zeta_2) = p(\zeta)$, and the second-order probability density function $p(\zeta_1, \zeta_2)$ only depends on the distance $\xi = x_1 - x_2$, known as correlation width. As a consequence, all random variables and, thus, the whole process have the same mean and the same mean square value, respectively. A stationary random process can always be centralized, so that its mean value becomes zero. Thus, for stationary processes, it follows from Equations 29.33 through 29.37

$$m_\zeta(x) = \text{const} = 0, \tag{29.38}$$

$$R_\zeta(x_1, x_2) = P_\zeta(x_1, x_2) = R_\zeta(\xi = x_1 - x_2), \tag{29.39}$$

$$m_{\zeta^2}(x) = \sigma_\zeta^2(x) = R_\zeta(\xi = 0) = \text{const}. \tag{29.40}$$

For centered processes, correlation and covariance are identical. According to Equation 29.39, the correlation depends on the correlation width ξ only; thus, it is called *correlation function* and characterizes the entire random process.

A random process is called *normal distributed* or a *Gaussian process* if for every value x_j, the corresponding random variable $\zeta_j = \zeta(x_j)$ has a probability density function

$$p(\zeta_j) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left[\frac{(\zeta_j - m_j)^2}{2\sigma_j^2}\right], \tag{29.41}$$

where $m_j = m_\zeta(x_j)$ and $\sigma_j = \sigma_\zeta(x_j)$. A normal distributed process can uniquely be defined by its mean and variance, that is, by its first two moments. The higher order moments can be expressed by these two quantities. For stationary, normal distributed processes, Equation 29.41 can be simplified using Equations 29.38 and 29.40. The graph of the probability density function (29.41) has the well-known bell shape (see Figure 29.7b). This curve shows a maximum value of $p_{\max} = 1/\left(\sigma_j\sqrt{2\pi}\right)$ at $\zeta_j = m_j$ and is symmetrical with respect to $\zeta_j = m_j$, with points of inflection at $\zeta_j = m_j \pm \sigma_j = \pm\zeta^*$, where $p(\pm\zeta^*) = p_{\max}/\sqrt{e}$ holds. Small values of the standard deviation σ_j mean that the quantity ζ_j is concentrated around m_j. From Equation 29.31, the probability can be calculated, so that the normal distributed random variable ζ_j lies in the interval of length $2k\,\sigma_j, k = 1, 2, \ldots$, around the mean m_j. These intervals are called confidential intervals or $k\sigma$ limits, $k = 1, 2, \ldots$:

$$\begin{aligned}
k = 1: \quad & \Pr(m - \sigma \le \zeta \le m + \sigma) = \Pr\left(|\zeta - m| \le \sigma\right) = 0.6827, \\
k = 2: \quad & \Pr\left(|\zeta - m| \le 2\sigma\right) = 0.9545, \\
k = 3: \quad & \Pr\left(|\zeta - m| \le 3\sigma\right) = 0.9973, \\
k = 3: \quad & \Pr\left(|\zeta - m| \le 4\sigma\right) = 0.99994.
\end{aligned} \tag{29.42}$$

Here, the index j has been dropped. The probability that a Gaussian random variable lies outside, for example, the 2σ limit is only $\Pr = 0,0455 \mathrel{\hat{=}} 4,55\%$.

A stationary random process is called *ergodic* if the ensemble averages of the random variables $\zeta_j = \zeta(x_j)$ is equal to the corresponding spacial (or temporal) averages taken along the coordinate x (or t) of any sufficiently long representative sample function $\zeta^{(r)}(x)$. This means in short

- Ensemble average is equal to sample average

For ergodic processes, it follows:

$$m_\zeta = \lim_{X\to\infty} \frac{1}{2X} \int_X^X \zeta^{(r)}(x)\,\mathrm{d}x, \tag{29.43}$$

$$R_\zeta(\xi) = \lim_{X\to\infty} \frac{1}{2X} \int_{-X}^{X} \zeta^{(r)}(x)\,\zeta^{(r)}(x-\xi)\mathrm{d}x, \tag{29.44}$$

$$m_{\zeta^2} = R_\zeta(\xi = 0) = \lim_{X\to\infty} \frac{1}{2X} \int_{-X}^{X} \left[\zeta^{(r)}(x)\right]^2 \mathrm{d}x, \tag{29.45}$$

$$P_\zeta(\xi) = \lim_{X\to\infty} \frac{1}{2X} \int_{-X}^{X} \left[\zeta^{(r)}(x) - m_\zeta\right]\left[\zeta^{(r)}(x-\xi) - m_\zeta\right]\mathrm{d}x, \tag{29.46}$$

$$\sigma_\zeta^2 = P_\zeta(\xi = 0) = \lim_{X\to\infty} \frac{1}{2X} \int_{-X}^{X} \left[\zeta^{(r)}(x) - m_\zeta\right]^2 \mathrm{d}x. \tag{29.47}$$

Often, it is difficult to decide if a stationary random process is ergodic or not. In the present case of guideway irregularities, ergodicity is at least plausible if one considers the generation of Figure 29.7: With respect to the information that can be gained, it is equivalent to analyzing a single very long measurement record (sample function) of, for example, highway surface irregularities or if one takes the ensemble of pieces of this record and performs a statistical analysis. In general, the averaging of sample functions is less time consuming than the ensemble averaging. However, in practical applications, neither an infinitely large ensemble nor a sample function of infinite length is given. So, both procedures result in approximations for the averages only.

In the following, always centered random processes with mean $m_\zeta = 0$ are assumed. Thus, there is no difference between mean square and variance or between correlation function and covariance function.

In technical applications, stationary processes are often characterized by their PSD or shortly spectral density $S_\zeta(\Omega)$. It can be gained from the correlation function $R_\zeta(\xi)$, by a Fourier transform,

$$S_\zeta(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_\zeta(\xi)\,\mathrm{e}^{-i\Omega\xi}\,\mathrm{d}\xi, \tag{29.48}$$

$$R_\zeta(\xi) = \int_{-\infty}^{\infty} S_\zeta(\Omega)\,\mathrm{e}^{i\Omega\xi}\,\mathrm{d}\Omega, \tag{29.49}$$

$$m_\zeta = 0 : P_\zeta = \sigma_\zeta^2 = R_\zeta(\xi = 0) = \int_{-\infty}^{\infty} S_\zeta(\Omega)\mathrm{d}\Omega. \tag{29.50}$$

Here, Ω with unit rad/m denotes a spacial circular frequency (wave number) that is connected to the wave length λ and the spacial frequency F by $\Omega = 2\pi/\lambda = 2\pi F$. Equations 29.48 and 29.49 are also known as Wiener–Chintschin relation; however, in literature, the factor $1/2\pi$ can also be found in front of the integral in Equation 29.49 or as factors $1/\sqrt{2\pi}$ in both equations. The benefit of the notation chosen here is clear from Equation 29.50, where the variance is equal to the spectral power. The random process ζ of guideway irregularities can equivalently be characterized by the correlation function $R_\zeta(\xi)$ in the spacial domain or by the spectral density $S_\zeta(\Omega)$ in the frequency domain. The variance σ_ζ^2 follows from $R_\zeta(\xi)$ for $\xi = 0$ or as an integral over the variance shares $S_\zeta(\Omega)\mathrm{d}\Omega$ related to the frequency Ω. Such a share can also be interpreted as variance after the process has passed a narrow band filter of bandwidth dΩ. Both functions $R_\zeta(\xi)$ and $S_\zeta(\Omega)$ are even functions,

$$R_\zeta(\xi) = R_\zeta(-\xi), \quad S_\zeta(\Omega) = S_\zeta(-\Omega). \tag{29.51}$$

This property allows an easy computation of the integrals in Equations 29.44 and 29.48 through 29.50, and leads to the introduction of the single-sided spectral density $\Phi_\zeta(\Omega)$ with nonnegative circular frequencies

$$\Phi_\zeta(\Omega) = \begin{cases} 2S(\Omega) & \Omega \geq 0 \\ \text{for} & . \\ 0 & \Omega < 0 \end{cases} \tag{29.52}$$

Correspondingly, $S(\Omega)$ is sometimes denoted as double-sided spectral density. Table 29.1 shows some often-used correlation functions $R_\zeta(\xi)$ and their corresponding single-sided spectral densities $\Phi_\zeta(\Omega)$.

Until now, only a scalar random process has been considered. If two scalar processes ζ_l, r are given, for example, the irregularities of the left and right track of an automobile, then the auto and cross-correlation functions and their corresponding spectra have to be distinguished. This is usually being done by double indices of the cross-correlation functions, where the Fourier transform is defined analogously to Equations 29.48 and 29.49,

$$S_{lr}(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{lr}(\xi)\, \mathrm{e}^{-i\Omega\xi}\, \mathrm{d}\,\xi \tag{29.53}$$

$$R_{lr}(\xi) = \int_{-\infty}^{\infty} S_{lr}(\Omega)\, \mathrm{e}^{i\Omega\xi}\, \mathrm{d}\,\Omega. \tag{29.54}$$

The cross-spectral densities are in general complex functions, where the following symmetries hold,

$$R_{lr}(\xi) = R_{rl}(-\xi), \quad S_{lr}(\Omega) = S_{rl}(-\Omega). \tag{29.55}$$

TABLE 29.1
Correlation Functions $R\zeta(\xi)$ and Corresponding Single-Sided Spectral Densities $\Phi_\zeta(\Omega)$ (All Constants Are Positive)

No.	Correlation Function $R(\xi) = \int_0^\infty \Phi(\Omega)\cos\Omega\xi d\Omega$	Single-Sided Spectral Density $\Phi(\Omega) = \frac{2}{\pi}\int_0^\infty R(\xi)\cos\Omega\xi d\xi$
I	$\sigma^2 e^{-\alpha\lvert\xi\rvert}$	$\frac{2\alpha\sigma^2}{\pi}\cdot\frac{1}{\Omega^2+\alpha^2}$
II	$\sigma^2 e^{-\alpha\lvert\xi\rvert}\cos\beta\xi$	$\frac{2\alpha\sigma^2}{\pi}\cdot\frac{\Omega^2+\alpha^2+\beta^2}{(\Omega^2-\alpha^2-\beta^2)^2+4\alpha^2\Omega^2}$
III	$\sigma^2 e^{-\alpha\lvert\xi\rvert}\left(\cos\beta\xi-\frac{\alpha}{\beta}\sin\beta\lvert\xi\rvert\right)$	$\frac{4\alpha\sigma^2}{\pi}\cdot\frac{\Omega^2}{(\Omega^2-\alpha^2-\beta^2)^2+4\alpha^2\Omega^2}$
IV	$\sigma^2 e^{-\alpha\lvert\xi\rvert}\left(\cos\beta\xi+\frac{\alpha}{\beta}\sin\beta\lvert\xi\rvert\right)$	$\frac{4\alpha\sigma^2}{\pi}\cdot\frac{\alpha^2+\beta^2}{(\Omega^2-\alpha^2-\beta^2)^2+4\alpha^2\Omega^2}$
	$\Omega_d = \sqrt{1-D^2}\cdot\Omega_0$	
V	$\sigma^2 e^{-D\Omega_0\lvert\xi\rvert}\left(\cos\Omega_d\xi+\frac{D\Omega_0}{\Omega_d}\sin\Omega_d\lvert\xi\rvert\right)$	$\frac{4D\Omega_0\sigma^2}{\pi}\cdot\frac{\Omega_0^2}{\left(\Omega^2-\Omega_0^2\right)^2+4D^2\Omega_0^2\Omega^2}$
VI	$\frac{\sigma^2}{\beta-\alpha}\left[\beta e^{-\alpha\lvert\xi\rvert}-\alpha e^{-\beta\lvert\xi\rvert}\right]$	$\frac{2\alpha\beta(\alpha+\beta)\sigma^2}{\pi}\cdot\frac{1}{(\Omega^2+\alpha^2)(\Omega^2+\beta^2)}$
VII	$\sigma^2 e^{-\alpha\xi^2}$	$\frac{\sigma^2}{\sqrt{\pi\alpha}}\cdot e^{-\frac{\Omega^2}{4\alpha}}$
VIII	$\sigma^2 e^{-\alpha\xi^2}\cos\beta\xi$	$\frac{\sigma^2}{\sqrt{4\pi\alpha}}\left[e^{-\frac{(\Omega+\beta)^2}{4\alpha}}+e^{-\frac{(\Omega-\beta)^2}{4\alpha}}\right]$
Harmonic function		
IX	$\sigma^2\cos\beta\xi$	$\sigma^2\delta(\Omega-\beta)$
White noise		
X	$q_w\delta(\xi)$	$\frac{1}{\pi}q_w$
Band limited white noise		
XI	$q_w\frac{\sin\alpha\xi}{\pi\xi}$	$\frac{1}{\pi}q_w$ if $0 \le \Omega \le \alpha$ 0 else
Ideal band pass		
XII	$2q_w\frac{\sin\alpha\xi/2\cdot\cos\beta\xi}{\pi\xi}$	$\frac{1}{\pi}q_w$ if $0 \le \beta-\frac{\alpha}{2} \le \Omega \le \beta+\frac{\alpha}{2}$ 0 else

A normalized measure for the correlation between two processes is given by the coherence function $\gamma(\Omega)$,

$$\gamma^2(\Omega) = \frac{\left|S_{lr}(\Omega)\right|^2}{S_l(\Omega)S_r(\Omega)}, 0 \le \gamma \le 1. \tag{29.56}$$

For $\gamma \equiv 0$, both processes, for example, the irregularities of the left and right track, are completely uncorrelated, and for $\gamma \equiv 1$, they are completely correlated. In general, the correlation is frequency dependent, as can be seen from Equation 29.56.

If n scalar processes ζ_ν, $\nu = 1(1)n$, are given, then it is suitable to combine them to a $n \times 1$-vector process $\boldsymbol{\zeta}$. The mean m_ζ is then replaced by the $n \times 1$-mean vector $\boldsymbol{m}_\zeta$ and the variance P_ζ by the $n \times n$-covariance matrix $\boldsymbol{P}_\zeta$,

$$\boldsymbol{m}_\zeta(x) = E\left\{\boldsymbol{\zeta}(x)\right\},$$

$$\boldsymbol{P}_\zeta(x) = E\left\{\left[\boldsymbol{\zeta}(x) - \boldsymbol{m}_\zeta(x)\right]\left[\boldsymbol{\zeta}(x) - \boldsymbol{m}_\zeta(x)\right]^{\mathrm{T}}\right\} = \boldsymbol{P}_\zeta^{\mathrm{T}}(x). \tag{29.57}$$

The variances of the single scalar processes appear as diagonal elements of the matrix $\boldsymbol{P}_\zeta$. It is clear that the variances are very important quantities to characterize the considered stationary, normal distributed, and ergodic processes. The variances can be interpreted in different ways as follows:

- Central second-order moment of a random variable, (29.35) and (29.57)
- Mean square of a centered sample function, (29.47)
- Spectral power of a centered process, (29.50)
- Square of the standard deviation of the probability density function, and, thus, defining the Gaussian bell-shaped distribution (29.41) and the size of the confidence intervals (29.42)

Example 29.3

White and coloured noise.

The correlation functions no. X, I, and IV of Table 29.1 should be taken to derive and interpret the corresponding PSDs.

Solution:

First, the Fourier transform for the single-sided PSD $\Phi(\Omega)$ will be derived. Introducing Euler's formula $\exp(\pm i\Omega\xi) = \cos\Omega\xi \pm i\sin\Omega t$ into Equations 29.48 and 29.49, and splitting the range of integration yields for Equation 29.48

$$S(\Omega) = \frac{1}{2\pi}\left[\int_{-\infty}^{0} R(\xi)(\cos\Omega\xi - i\sin\Omega\xi)\,\mathrm{d}\xi + \int_{0}^{\infty} R(\xi)(\cos\Omega\xi - i\sin\Omega\xi)\,\mathrm{d}\xi\right]. \tag{29.58}$$

Using $\xi = -\bar{\xi}$ in the first integral, where $\bar{\xi}$ is a new variable, and regarding $R(\xi) = R(-\xi) = R(\bar{\xi})$ according to (29.51) results in

$$2S(\Omega) = \frac{1}{\pi}\left[\int_{0}^{\infty} R(\bar{\xi})(\cos\Omega\bar{\xi} + i\sin\Omega\bar{\xi})\,\mathrm{d}\bar{\xi} + \int_{0}^{\infty} R(\xi)(\cos\Omega\xi - i\sin\Omega\xi)\,\mathrm{d}\xi\right]. \tag{29.59}$$

Putting the right-hand side together and regarding $2S(\Omega) = \Phi(\Omega)$, it follows

$$\Phi(\Omega) = \frac{2}{\pi}\int_0^\infty R(\xi)\cos\Omega\xi\,d\xi. \tag{29.60}$$

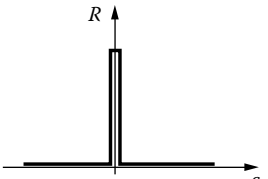
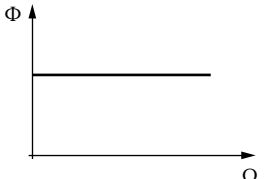
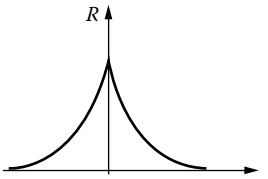
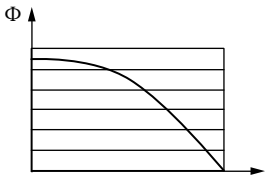
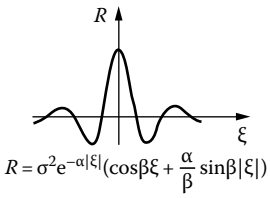
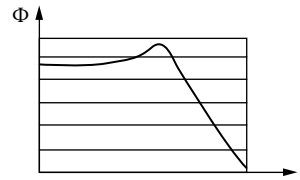
The inverse transform follows in the same manner,

$$R(\xi) = \int_0^\infty \Phi(\Omega)\cos\Omega\xi\,d\Omega. \tag{29.61}$$

Introducing the three correlation functions from Table 29.1 into Equation 29.61 yields the following results that are visualized in Table 29.2.

(1) Correlation function no. $X, R(\xi) = q_w\delta(\xi)$:

TABLE 29.2
Comparison of Correlation Function $R_\xi(\xi)$ and Spectral Density $\Phi_\xi(\Omega)$ of White and Colored Noise

Denomination	Correlation Function $R(\xi)$	Spectral Density $\Phi(\Omega)$ (Double Logarithmic Scale)
White noise	 $R = q_w\delta(\xi)$	 $\Phi = \frac{1}{\pi} q_w$
Colored noise I	 $R = \sigma^2 e^{-\alpha\lvert\xi\rvert}$	 $\Phi = \frac{2\alpha\sigma^2}{\pi} \times \frac{1}{\alpha^2 + \Omega^2}$
Colored noise IV	 $R = \sigma^2 e^{-\alpha\lvert\xi\rvert}(\cos\beta\xi + \frac{\alpha}{\beta}\sin\beta\lvert\xi\rvert)$	 $\Phi = \frac{4\alpha\sigma^2}{\pi} \times \frac{\alpha^2 + \beta^2}{(\alpha^2 - \beta^2 + \Omega^2)^2 + 4\alpha^2\beta^2}$

The evaluation of Equation 29.61 yields

$$\frac{2}{\pi}\Phi(\Omega) = \int_0^\infty q_w \delta(\xi)\cos\Omega\xi \mathrm{d}\xi = \frac{q_w}{\pi} = \text{const.} \tag{29.62}$$

For the computation of the integral in Equation 29.62, it has to be considered that the lower limit of the integral coincides with the coordinate where the Dirac distribution $\delta(\xi)$ is nonzero, thus, not the entire integrand but only one half of it is filtered out by the integration. The same results follow, if the transform (29.48) is applied, where $S(\Omega) = \Phi(\Omega)/2$ has to be introduced. As a consequence, the random process no. X that is completely uncorrelated for $\xi \neq 0$ has a constant PSD for all frequencies. Such a process is called white noise $w(x)$, because it contains a constant spectrum like white light. However, this noise process has an infinite variance,

$$P_w = \sigma_w^2 = R_w(\xi = 0) = \int_0^\infty \Phi(\Omega)\mathrm{d}\Omega = \infty. \tag{29.63}$$

This shows that white noise is a mathematical idealization and does not occur in nature. However, it is a useful limit case of real noise and it can mathematically be handled easily; thus, it is applied very often. Stationary Gaussian white noise with zero mean, $m_w = 0$, and intensity q_w will be denoted by

$$w(x) \sim N(0, q_w). \tag{29.64}$$

(2) Correlation function no. I, $R(\xi) = \sigma^2 \mathrm{e}^{-\alpha|\xi|}, \alpha > 0$:

The solution requires the evaluation of Equation 29.60,

$$\Phi(\Omega) = \frac{2\sigma^2}{\pi}\int_0^\infty \mathrm{e}^{-\alpha\xi}\cos\Omega\xi\,\mathrm{d}\xi = \frac{2\sigma^2}{\pi}\left[\frac{\mathrm{e}^{-\alpha\xi}}{\Omega^2+\alpha^2}(-\alpha\cos\Omega\xi + \Omega\sin\Omega\xi)\right]_{\xi=0}^{\xi=\infty},$$

$$\Phi(\Omega) = \frac{2\sigma^2}{\pi}\frac{\alpha}{\Omega^2+\alpha^2}. \tag{29.65}$$

Here, the computation is very simple, since $|\xi| = \xi$ for $\xi \geq 0$ holds and the right-hand side of Equation 29.65 vanishes at the upper limit. The result is a coloured noise process with a maximum PSD of $\Phi_{max} = 2\sigma^2/(\pi\alpha)$ at the spacial frequency $\Omega = 0$, and it follows an asymptotic behavior $\Omega \sim 1/\Omega^2$ for $\Omega \to \infty$. Here, the variance $\sigma^2 = R(\xi = 0)$ is finite. The correlation decreases with increasing distance ξ monotonically. Table 29.2 shows clearly the differences to a white noise process.

(3) Correlation function no. IV, $R(\xi) = \sigma^2 \mathrm{e}^{-\alpha|\xi|}\left(\cos\beta\xi + \frac{\alpha}{\beta}\sin\beta|\xi|\right), \alpha > 0$:

In this case, the solution is analogous to (29.59) regarding trigonometrical transformations for products of harmonic functions,

$$\Phi(\Omega) = \frac{2\sigma^2}{\pi}\int_0^\infty \mathrm{e}^{-\alpha\xi}\left(\cos\beta\xi + \frac{\alpha}{\beta}\sin\beta\xi\right)\cos\Omega\xi\,\mathrm{d}\xi$$

$$= \frac{2\sigma^2}{\pi}\int_0^\infty \left\{\mathrm{e}^{-\alpha\xi}\frac{1}{2}\left[\cos(\beta+\Omega)\xi + \cos(\beta-\Omega)\xi\right] + \mathrm{e}^{-\alpha\xi}\frac{\alpha}{2\beta}\left[\sin(\beta+\Omega)\xi + \sin(\beta-\Omega)\xi\right]\right\}\mathrm{d}\xi$$

$$\begin{aligned}\Phi(\Omega) &= \left\{ e^{-\alpha\xi}\left[\frac{1}{\alpha^2+(\beta+\Omega)^2}\left(-\alpha\cos(\beta+\Omega)\xi+(\beta+\Omega)\sin(\beta+\Omega)\xi\right)\right.\right.\\ &+\frac{1}{\alpha^2+(\beta-\Omega)^2}\left(-\alpha\cos(\beta-\Omega)\xi+(\beta-\Omega)\sin(\beta-\Omega)\xi\right)\\ &+\frac{\alpha}{\beta\left(\alpha^2+(\beta+\Omega)^2\right)}\left(-\alpha\sin(\beta+\Omega)\xi-(\beta+\Omega)\cos(\beta+\Omega)\xi\right)\\ &\left.\left.+\frac{\alpha}{\beta\left(\alpha^2+(\beta-\Omega)^2\right)}(-\alpha\sin(\beta-\Omega)\xi-(\beta-\Omega)\cos(\beta-\Omega)\xi)\right]\right\}_{\xi=0}^{\xi=\infty}\\ &=\frac{\sigma^2}{\pi}\left[\frac{\alpha}{\alpha^2+(\beta+\Omega)^2}+\frac{\alpha}{\alpha^2+(\beta-\Omega)^2}\right.\\ &\left.+\frac{\alpha}{\beta}\left(\frac{\beta+\Omega}{\alpha^2+(\beta+\Omega)^2}+\frac{\beta-\Omega}{\alpha^2+(\beta-\Omega)^2}\right)\right]\\ &=\frac{4\alpha\sigma^2}{\pi}\frac{\alpha^2+\beta^2}{N}. \end{aligned} \tag{29.66}$$

In the last expression, the denominator can be expressed in a different manner,

$$\begin{aligned}N &= \left[\alpha^2+(\beta+\Omega)^2\right]\left[\alpha^2+(\beta-\Omega)^2\right]\\ &=\left(\Omega^2-(\alpha^2-\beta^2)^2\right)+4\alpha^2\Omega^2=(\Omega^2+\alpha^2-\beta^2)^2+4\alpha^2\beta^2. \end{aligned} \tag{29.67}$$

Again, the result is a colored noise process with the following maximum PSD Φ_{max},

$$\Phi_{max}=\begin{cases}\dfrac{\alpha\sigma^2}{\pi} & \dfrac{\alpha^2+\beta^2}{\alpha^2\beta^2} \quad \text{at} \quad \Omega=\sqrt{\beta^2-\alpha^2} \quad \text{for} \quad \beta\geq\alpha\\ \dfrac{\alpha\sigma^2}{\pi} & \dfrac{4}{\alpha^2+\beta^2} \quad \text{at} \quad \Omega=0 \quad \text{for} \quad \beta\leq\alpha\end{cases}. \tag{29.68}$$

It results in asymptotic behavior $\Phi \sim 1/\Omega^4$ for $\Omega \to \infty$. The variance $\sigma^2 = R(\xi = 0)$ is finite. Here, the correlation decreases with increasing distance ξ in an oscillating manner around zero, Table 29.2.

29.2.2 Models of Road Roughness Profiles

The evaluation and preparation of numerous measurement data of road roughness profiles show a tendency toward standardization. A simple but often-used road roughness model reads

$$\Phi_\zeta(\Omega)=\Phi_0\left(\frac{\Omega_0}{\Omega}\right)^w, \quad 0<\Omega_I\leq\Omega\leq\Omega_\Pi<\infty, \tag{29.69}$$

where the frequency range is bounded. Here, Ω_0[rad/m] denotes a standardized spacial circular frequency (wave number), $\Phi_0 = \Phi_\zeta(\Omega_0)$[m^2/(rad/m)] characterizes the roughness at Ω_0 and can be

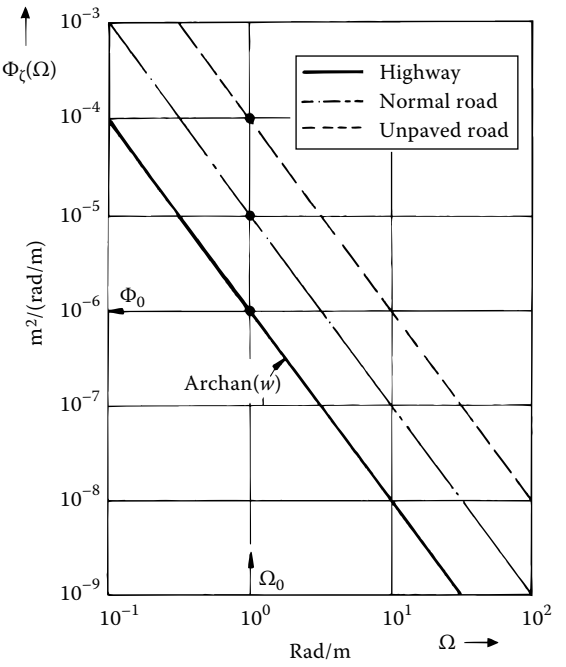


FIGURE 29.8 Simple road roughness models with approximate values of the roughness level $\Phi_0 = \Phi(\Omega_o)$, $\Omega_0 = 1$ rad/m.

considered as roughness level, and w is called waviness and characterizes whether the road contains long wave lengths (w large) or short wave lengths (w small) with pronounced spectral densities. For the wavelength λ, the relation $\lambda = 2\pi/\Omega$ holds. The waviness w of road profiles lies in the range $1.75 \le w \le 2.25$ with $w \approx 2$ as an average value. Equation 29.69 is usually plotted in a Ω, Φ–diagram with double logarithmic scale and results in decreasing straight lines with slopes ($-w$) (Figure 29.8).

A similar but more complicated model reads, [13],

$$\Phi_\zeta(\Omega) = \begin{cases} \Phi_0 \left(\dfrac{\Omega_0}{\Omega} \right)^{w_1} & 0 < \Omega_\mathrm{I} \le \Omega \le \Omega_0, \\ & \text{for} \\ \Phi_0 \left(\dfrac{\Omega_0}{\Omega} \right)^{w_2} & \Omega_0 \le \Omega \le \Omega_\mathrm{II} < \infty, \end{cases} \tag{29.70}$$

where different values w_1, w_2, $w_1 < w_2$, of the waviness occur in the two frequency ranges.

The road roughness models (29.68) and (29.70) are approximations to measured PSD in a distinct frequency range $0 < \Omega_\mathrm{I} \le \Omega \le \Omega_\mathrm{II} < \infty$. In the limit case $\Omega \to 0$, both models result in infinite values of the PSD, $\Phi_\zeta(0) \to \infty$, and, thus, in infinite variances, $P_\zeta \to \infty$. To avoid this unrealistic case, extended road roughness models have been suggested that hold in the entire frequency range. These models are colored noise processes characterized by the PSD

$$\mathrm{I}: \quad \Phi_{\zeta_\mathrm{I}}(\Omega) = \frac{2\alpha\sigma^2}{\pi} \frac{1}{\alpha^2 + \Omega^2}, \tag{29.71}$$

$$0 \le \Omega < \infty,$$

$$\text{II}: \Phi_{\zeta\text{II}}(\Omega) = \frac{2\alpha\sigma^2}{\pi} \frac{\Omega^2 + \alpha^2 + \beta^2}{(\Omega^2 - \alpha^2 - \beta^2)^2 + 4\alpha^2\Omega^2}, \tag{29.72}$$

where α, β, and σ^2 are positive constants. The corresponding correlation functions can be calculated using

$$R_\zeta(\xi) = \int_0^\infty \Phi_\zeta(\Omega) \cos\Omega_\xi \, \mathrm{d}\Omega, \tag{29.73}$$

Inserting (29.71) and (29.72) into (29.73) results in

$$\text{I}: \; R_{\zeta\text{I}}(\xi) = \sigma^2 \, \mathrm{e}^{-\alpha|\xi|}, \tag{29.74}$$

$$\text{II}: \quad R_{\zeta\text{II}}(\xi) = \sigma^2 \, \mathrm{e}^{-\alpha|\xi|} \cos\beta\xi, \tag{29.75}$$

see also Table 29.2, where $\sigma^2 = R_\zeta(\xi = 0)$ denotes the finite variance of the road roughness profile. For $\Omega \to \infty$, both PSDs (29.71) and (29.72) decay according to $\Phi \sim 1/\Omega^2$. Thus, they approximate measured road roughness profiles very well.

29.2.3 Models of Vehicle Excitation

From the models of the road roughness profiles $\zeta(x)$ shown in Section 29.2, corresponding models of the vehicle excitation $\zeta(t)$ in the time domain can be derived. Here, permanent road contact is assumed and, in a first step, only one contact point is considered. The transformation from space to time domain, $\zeta(x) \to \zeta(t)$, is performed considering the vehicle speed $v(t)$,

$$\mathrm{d}\, x(t) = v(t)\mathrm{d}t, \quad x(t) = x(t_0) + \int_{t_0}^{t} v(\tau)\, \mathrm{d}\,\tau. \tag{29.76}$$

From the road roughness profiles that are stationary in space domain, it generally follows non-stationary vehicle excitation processes in the time domain if the vehicle speed $v(t)$ is time varying [14,15]. This case will not be considered here. Throughout, we assume a constant vehicle speed, $v = \text{const}$, resulting in stationary vehicle excitation processes. For $t_0 = 0$ and $x(t_0) = 0$ from (29.76), it follows

$$x(t) \; = vt, \quad \xi = v\tau, \quad \omega = v\Omega, \quad v = \text{const}. \tag{29.77}$$

Here, τ means the correlation time corresponding to the correlation width ξ and ω with unit rad/s denotes the temporal circular frequency. Since the variances $R_\zeta(0)$ of road profiles $\zeta(x)$ and vehicle excitations $\zeta(t)$ must be equal, Equation 29.73 yields $\Phi_\zeta(\omega)\mathrm{d}\omega = \Phi_\zeta(\Omega)\, \mathrm{d}\Omega$ that results in (29.77) in the PSD $\Phi_\zeta(\omega)$ [m^2/(rad/s)] of the vehicle excitation

$$\Phi_\zeta(\omega) = \frac{1}{v} \Phi_\zeta\left(\Omega = \frac{\omega}{v}\right). \tag{29.78}$$

For example, from the road roughness model (29.69) with the average waviness $w = 2$, it follows

$$\tilde{\Phi}_\zeta(\omega) = \frac{1}{v}\tilde{\Phi}_0\left(\frac{v\Omega_0}{\omega}\right)^2 = v\,\tilde{\Phi}_0\left(\frac{\Omega_0}{\omega}\right)^2. \tag{29.79}$$

Until now, we considered stochastic excitation processes $\zeta(t)$ of the vehicle displacements only. But the differentiated processes $\dot{\zeta}(t)$ and $\ddot{\zeta}(t)$ of vehicle velocity and vehicle acceleration, respectively, can also be regarded. The corresponding PSDs read, [11],

$$\Phi_{\dot{\zeta}}(\omega) = \omega^2\Phi_\zeta(\omega), \quad \Phi_{\ddot{\zeta}}(\omega) = \omega^4\Phi_\zeta(\omega). \tag{29.80}$$

Equation 29.80 applied to (29.79), for example, yields a white noise process for the vehicle excitation velocity $\dot{\zeta}(t)$ characterized by the PSD

$$\tilde{\Phi}_{\dot{\zeta}}(\omega) = \omega^2 v\,\tilde{\Phi}_0\left(\frac{\Omega_0}{\omega}\right)^2 = v\,\tilde{\Phi}_0\Omega_0^2 = \text{const}. \tag{29.81}$$

The corresponding correlation function reads

$$R_{\dot{\zeta}}(\tau) = q_{\dot{\zeta}}\delta(\tau), \quad q_{\dot{\zeta}} = \pi v\,\tilde{\Phi}_0\Omega_0^2, \tag{29.82}$$

where $q_{\dot{\zeta}}$ denotes the noise intensity of $\dot{\zeta}(t)$ and $\delta(\tau)$ is the Dirac distribution. Assuming the waviness $w = 4$ from Equation 29.69, similarly, a white noise process for the vehicle excitation acceleration $\ddot{\zeta}(t)$ follows,

$$\tilde{\tilde{\Phi}}_{\ddot{\zeta}}(\omega) = \omega^4 v\tilde{\tilde{\Phi}}_0\left(\frac{\Omega_0}{\omega}\right)^4 = v\,\tilde{\tilde{\Phi}}_0\Omega_0^4 = \text{const}, \tag{29.83}$$

$$R_{\ddot{\zeta}}(\omega) = q_{\ddot{\zeta}}\delta(\tau), \quad q_{\ddot{\zeta}} = \pi v\,\tilde{\tilde{\Phi}}_0\Omega_0^4. \tag{29.84}$$

The white noise processes are characterized by infinite variances and, thus, they are not realistic. On the other hand, assuming vehicle excitations like (29.81)–(29.84), the effort for the computation of vehicle random vibrations can be reduced considerably. Therefore, the white noise velocity excitation of road vehicles is of importance as a technical approximation.

More realistic models of vehicle excitations $\zeta(t)$ are based on Equations 29.71 through 29.75, that is, they are stationary Gaussian colored noise processes. These processes can be generated as a steady-state solution of a system of linear time-invariant ordinary differential equations (ODEs) that are excited by white noise $w(t)$. These ODEs change—graphically speaking—the shape of the PSD or corresponding correlation function; thus, they are called shape filters (Figure 29.9).

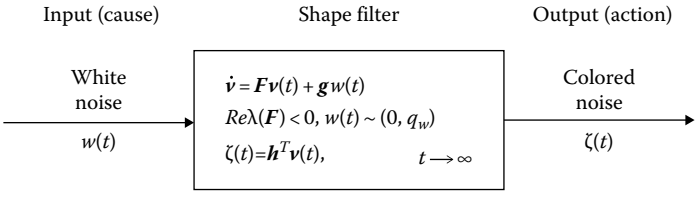


FIGURE 29.9 Visualization of a shape filter.

Generally holds, if a stationary, ergodic Gaussian process excites a system of linear time-invariant ODEs, then the properties stationarity, normal distribution, and ergodicity will be transferred from the system input to the system output. The shape filter eqs. read

$$\zeta(t) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{v}(t), \tag{29.85}$$

$$\dot{\boldsymbol{v}}(t) = \boldsymbol{F}\boldsymbol{v}(t) + \boldsymbol{g}w(t), \mathrm{Re}\,\lambda(F) < 0, w(t) \sim N(0, q_w). \tag{29.86}$$

In steady state, the colored noise process $\zeta(t)$ results by the superposition of the state variables $v_i(t)$, $i = 1(1)m$, that are combined to the $m \times 1$-state vector v, (29.85). This state vector v is solution of a system of ODEs that are combined to the shape filter state equation (29.86), where the system matrix F is asymptotically stable and the excitation is given by white noise $w(t)$ with zero mean and intensity q_w. The quantities $\boldsymbol{F}$, $\boldsymbol{g}$, and $\boldsymbol{h}$ characterize the shape filter completely. For the models (29.71)–(29.75), these quantities read

$$\mathrm{I}: \quad F = -\alpha v, \quad \boldsymbol{g} = g, \quad \boldsymbol{h} = 1, \quad (g^2 q_w = 2\alpha v \sigma^2), \tag{29.87}$$

$$\mathrm{II}: \quad F = \begin{bmatrix} 0 & 1 \\ -(\alpha^2 + \beta^2)v^2 & -2\alpha v \end{bmatrix}, \quad \boldsymbol{g} = g\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} v\sqrt{\alpha^2 + \beta^2} \\ 1 \end{bmatrix}, \quad (g^2 q_w = 2\alpha v \sigma^2). \tag{29.88}$$

Since the product $g^2 q_w$ of the shape filter quantities follows from a product of model parameters, there is a possibility of choice. If we put $g = 1$, for example, then for both shape filters, the intensity q_w of the white noise excitation reads $q_w = 2\,\alpha\, v\, \sigma^2$. This shows that the intensity of the vehicle excitation grows proportional to the vehicle speed v.

In general, the shape filter quantities should be determined directly from measured PSDs $\Phi_\zeta(\Omega)$ rather than from analytical approximations such as those shown here. However, in this case, the order m of the filter has to be chosen. For $m = 2$, this procedure is extensively described in [16].

Until now, only vehicle excitations caused by a single contact point have been considered resulting in a scalar stochastic excitation process. However, real multiaxle vehicles have multiple road contacts. For $i = 1(1)q$ axles and $k = 1(1)s$ tracks, there are $q{\cdot}s$ contact points resulting in correspondingly many scalar excitation processes that can be combined to a $(q{\cdot}s) \times 1 -$ vector process ζ of vehicle excitation.

In case of subsequent contact points in a single track, the excitation processes $\zeta_i(t)$, $i = 1(1)q$ can be gained from the excitation process $\zeta_1(t) \equiv \zeta(t)$ of the first contact point by a time shift t_i

$$\zeta_i(t) = \zeta(t - t_i), \quad t_i = \frac{l_i}{v}, \quad 0 = t_1 < t_2 < \cdots < t_q, \quad i = 1(1)q, \tag{29.89}$$

where

l_i denotes the distance between the first and the ith axle

v is the const is the vehicle speed

Further details can be found in [17].

In case of parallel tracks, the statistical relationship between the parallel-acting excitation processes have to be regarded. Measurements of correlation functions are reported in [9]. A corresponding shape filter is given by [14].

A typical vector process characterizing the excitation of a four-wheel vehicle reads ($f \triangleq$ front, $r \triangleq$ rear, $l \triangleq$ left, $r \triangleq$ right),

$$\boldsymbol{\zeta}(t,T) = \begin{bmatrix} \zeta_{lf}(t) \\ \zeta_{rf}(t) \\ \zeta_{lr}(t) \\ \zeta_{rr}(t) \end{bmatrix} = \begin{bmatrix} \zeta_{l}(t) \\ \zeta_{r}(t) \\ \zeta_{l}(t-T) \\ \zeta_{r}(t-T) \end{bmatrix}, \quad T \equiv t_2 = \frac{l_2}{v}, \tag{29.90}$$

where the delay time T follows from the axle distance l_2 and the vehicle speed v.

29.3 COMPUTATION OF THE RANDOM VEHICLE RESPONSE

The entire vehicle system modelled by the equation of motion (see Chapter 2) and excited by road roughness can be described by the non linear state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{a}(\boldsymbol{x}, \boldsymbol{w}, t), \tag{29.91}$$

where
- $\boldsymbol{x}$ is the state vector
- $\boldsymbol{w}$ characterizes the excitation
- $\boldsymbol{a}$ denotes a nonlinear vector function

In the following, we will assume that Equation 29.91 can be linearized resulting in the linear state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}(t)\,\boldsymbol{x}(t) + \boldsymbol{B}(t)\,\boldsymbol{w}(t), \tag{29.92}$$

where, in many cases, the state matrix $\boldsymbol{A}$ and the input matrix $\boldsymbol{B}$ are time invariant. The state equation serves immediately as the basis of evaluation of the driving stability (Section 29.1.1). However, in case of ride comfort evaluation, a special shape filter has to be considered. The input of this shape filter is the acceleration $a(t)$ at the seat position, which is composed of the states and excitation quantities of the entire vehicle system,

$$a(t) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}(t) + \boldsymbol{d}^{\mathrm{T}}\boldsymbol{w}(t), \tag{29.93}$$

where
- c is a weighting vector of the states
- d a weighting vector of the excitations

Thus, the state Equation 29.92 extended by the shape filter Equation 29.2 reads

$$\underbrace{\begin{bmatrix} \dot{\boldsymbol{x}} \\ \dot{\bar{\boldsymbol{v}}} \end{bmatrix}}_{\dot{\tilde{\boldsymbol{x}}}} = \underbrace{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{g}\boldsymbol{c}^{\mathrm{T}} & \bar{\boldsymbol{F}} \end{bmatrix}}_{\tilde{\boldsymbol{A}}} \underbrace{\begin{bmatrix} \boldsymbol{x} \\ \bar{\boldsymbol{v}} \end{bmatrix}}_{\tilde{\boldsymbol{x}}} + \underbrace{\begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{g}\boldsymbol{d}^{\mathrm{T}} \end{bmatrix}}_{\tilde{\boldsymbol{B}}} \boldsymbol{w}, \tag{29.94}$$

$$\dot{\tilde{\boldsymbol{x}}} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{B}}\boldsymbol{w}.$$

The extended system (29.94) has exactly the same structure as the vehicle system (29.92), so that we can consider system (29.92) subsequently without loss of generality. The required computational methods can be subdivided into numerical and analytical methods that will be described in the following sections.

29.3.1 Numerical Simulation

The given set of linear or nonlinear differential Equations 29.91 and 29.92 can be analyzed by numerical integration (see Chapters 2, 3, and 5). For random excitations, a single realization is sufficient due to ergodicity. In case of nonlinear systems, the random response in general is not normally distributed any more. Even so, in many cases, one computes the first two moments only for evaluation purposes. However, the effort of numerical simulations is always large due to the complex vehicle system models. Thus, a proper choice of the integration methods is very important. Unfortunately, it is not possible to give general recommendations, since on the one hand new integration methods will be developed in numerical mathematics, and on the other hand, the performance of computers is still increasing. Often, the users apply the integration methods at hand and perform test runs for comparison.

Example 29.4

Simulation of vertical motions of vehicles.

Different integration methods shall be applied to the simulation of the vertical motion of the front and rear axle of a nonlinear 16-DOF-vehicle model under a ramp excitation. Furthermore, a five-DOF-vehicle model is considered as a test example of a random excitation.

Solution:

The ramp is crossed first by the front axle and then by the rear axle; this results in a time delay in the corresponding vibration responses (Figure 29.10).
Similar to Rill [18], the following integration methods shall be compared:

- Single-step methods
  (Runge–Kutta–Fehlberg procedure of orders 5 and 6 [RKF5, RKF6]),
- Predictor-corrector multistep method
  (Shampine–Gordon procedure [SGDE]),
- Extrapolation method
  (Stoer–Bulirsch procedure [SB]).

The results of the large-vehicle model are shown in Figure 29.11. It can be seen that the Shampine–Gordon procedure [19] leads to the best results with respect to computing time and accuracy.

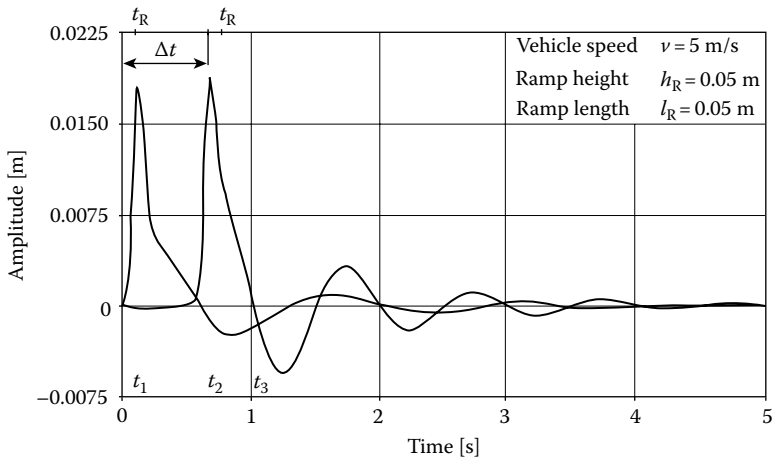


FIGURE 29.10 Vertical motion of the front and rear wheels of a nonlinear 16-DOF-vehicle model with ramp excitation.

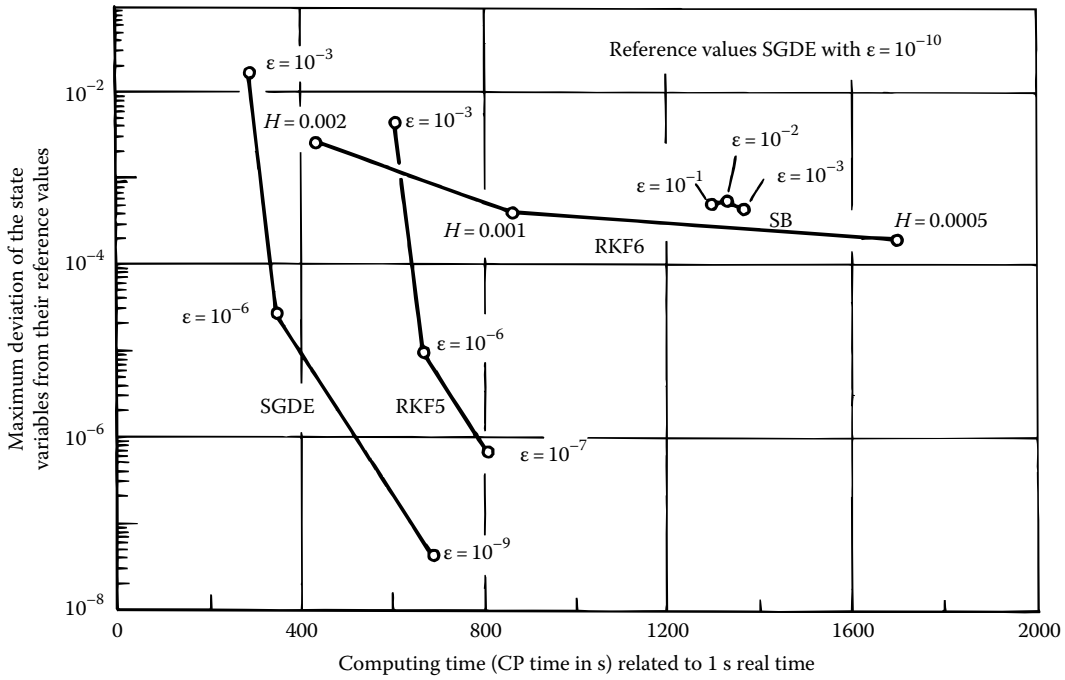


FIGURE 29.11 Comparison of integration methods for a 16-DOF-vehicle model with ramp excitation (error bound ε, step size H).

The random excitation of the small-vehicle model is generated by a random superposition of harmonic functions characterizing the road roughness profile (Figure 29.12).

The vertical motion of the front axle, see Figure 29.13, follows the road profile; however, the high-frequency components of the excitation are suppressed. Nevertheless, the integration procedure must take into account the high frequencies. Applying the procedures mentioned earlier lead to the results shown in Figure 29.14. Here, the simple Runge–Kutta–Fehlberg procedures are superior due to the fast evaluation of the right-hand side of the differential equations.

The numerical integration procedures result in time histories that have to be evaluated in order to assess ride comfort and safety. In case of random excitation, the computation of the performance criteria requires time averages that can be gained directly from spectral density or covariance analysis.

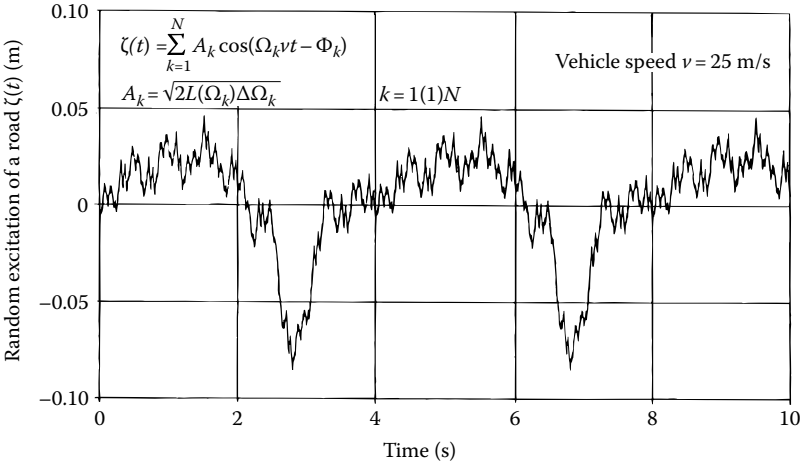


FIGURE 29.12 Realization of a random road excitation process.

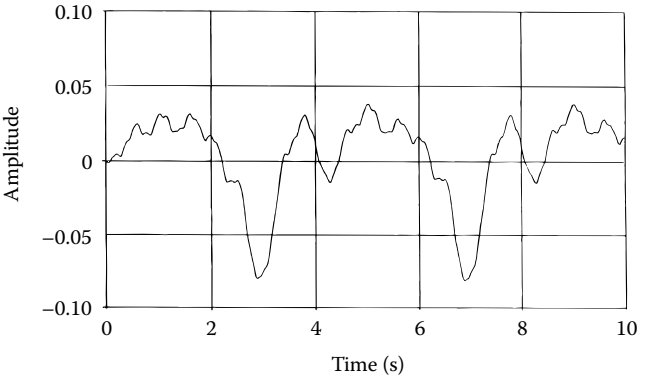


FIGURE 29.13 Vertical motion of the front wheel of a nonlinear five-DOF-vehicle model with random excitation.

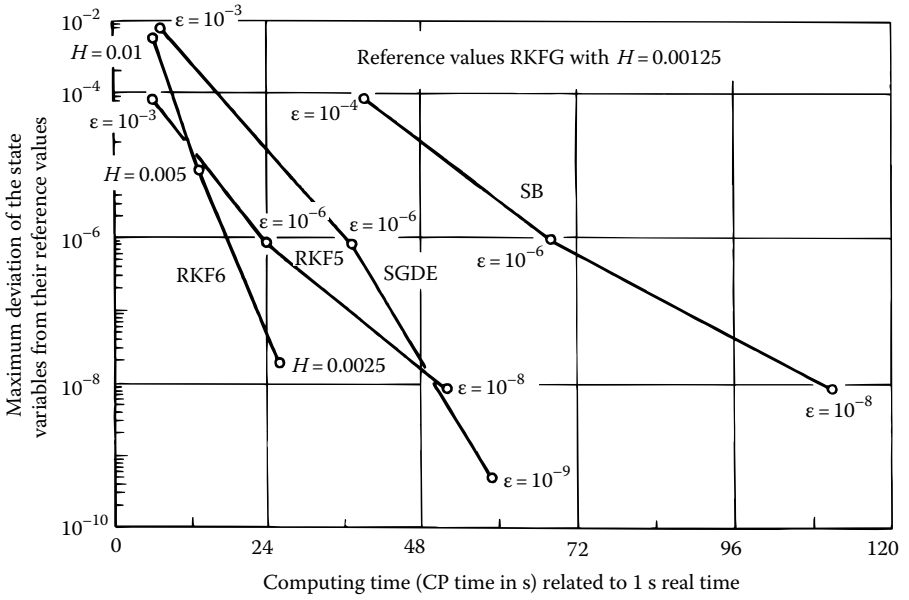


FIGURE 29.14 Comparison of integration methods for a five-DOF-vehicle model with random excitation (error bound ε, step size H).

29.3.2 Spectral Density Analysis

The investigation of random vibrations can be performed in the frequency domain using spectral analysis or in the time domain applying covariance analysis.

Assuming that $x(t) = 0$ is an asymptotically stable equilibrium position, then from $E\{\boldsymbol{w}(t)\} = \mathbf{0}$, it follows $E\{\boldsymbol{x}(t)\} = \mathbf{0}$, that is, the mean value of the state vector vanishes in the steady state. Thus, the essential goal is to calculate the characterizing variances, for example, the variance $\sigma_{\bar{a}}^2 = E\left\{\bar{a}^2(t)\right\}$ of the frequency weighted acceleration $\bar{a}(t)$.

First, we calculate the PSD that is a characteristic quantity of a random process in the frequency domain. In particular, for linear systems, the PSD of the input and output processes are related to each other by its FRFs. However, one should keep in mind that the ride comfort and safety criteria depend primarily on variances and not on PSDs.

The $n \times n$-PSD-matrix $\boldsymbol{S}_x(\omega)$ of the entire vehicle system described by the state Equation 29.92, where $\boldsymbol{A}$ = const and $\boldsymbol{B}$ = const has been assumed, follows from the white noise process $\boldsymbol{w}(t) \sim (\boldsymbol{0}, \boldsymbol{Q}_W)$ as

$$\boldsymbol{S}_x(\omega) = (i\omega\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{B}\,\boldsymbol{Q}_W\boldsymbol{B}^{\mathrm{T}}(-i\omega\boldsymbol{E} - \boldsymbol{A})^{-\mathrm{T}}. \tag{29.95}$$

Thus, the PSD-matrix $\boldsymbol{S}_x(\omega)$ is gained from the input intensity matrix $\boldsymbol{Q}_W$ by matrix multiplications with the FRF-matrix $\boldsymbol{F}_x = (i\omega\boldsymbol{E} - \boldsymbol{A})^{-1}$. Since the state vector $x(t)$ and the system matrix A often can be partitioned, it is possible to obtain intermediate results for the PSD matrices of the road excitation, the suspension system, and the chassis, respectively. This reduces possibly the numerical amount considerably.

The scalar PSD $S_a(\omega)$ of the scalar acceleration $a(t)$, given by Equation 29.93, reads

$$S_a(\omega) = (\boldsymbol{c}^{\mathrm{T}}\boldsymbol{F}_x(\omega)\boldsymbol{B} + \boldsymbol{d}^{\mathrm{T}})\,\boldsymbol{Q}_W\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{F}_x^{\mathrm{T}}(-\omega)\,\boldsymbol{c} + \boldsymbol{d}\right), \tag{29.96}$$

where the $n \times n$-FRF-matrix $\boldsymbol{F}_x$ has been used again. Equation 29.95 shows that the PSD is a quadratic expression, that is, the PSD of a sum of quantities leads to additional terms in the result. Using the PSD $S_a(\omega)$, Equation 29.95, the frequency weighting due to Equation 29.16 can be performed very easily; it yields

$$S_{\bar{a}}(\omega) = \alpha^2 \left|F(\omega)\right|^2 S_a\,. \tag{29.97}$$

The variance follows from the PSD by integration over an infinite interval,

$$\sigma_{\bar{a}}^2 = \int_{-\infty}^{\infty} S_{\bar{a}}(\omega)\,\mathrm{d}\,\omega = \int_{0}^{\infty} \Phi_{\bar{a}}(\omega)\,\mathrm{d}\,\omega. \tag{29.98}$$

For comparison with Equation 29.16 in Equation 29.98, the single-sided PSD $\Phi_{\bar{a}}(\omega)$ with $\Phi_{\bar{a}}(\omega) = 2S_{\bar{a}}(\omega)$ has been introduced.

The numerical computation of the PSD (29.98) is generally not difficult. However, one has to deal with complex matrices, many approximation points, and a very large integration interval; this can lead to numerical errors and considerable computation times. The integration (29.98) can be avoided by applying the covariance analysis as shown in the next section.

Special care is required in the spectral analysis of multiaxle vehicles. The excitation at different axles is given by an excitation function with time delay,

$$\begin{aligned} \boldsymbol{B}\boldsymbol{w}(t) &= \sum_{i=1}^{m} \boldsymbol{B}_i\zeta_i(t), \\ \zeta_i(t) &= \zeta(t - t_i), \quad 0 = t_1 < t_2 < \cdots < t_m, \\ t_i &= \frac{l_i}{v}, \quad i = 1, \ldots, m, \end{aligned} \tag{29.99}$$

where

l_i denotes the distance between the front axle and the ith axle

v is the const is the vehicle speed

The PSD matrix of a two-axle vehicle under random excitation reads

$$\boldsymbol{S}_x(\omega) = (i\omega\boldsymbol{E} - \boldsymbol{A})^{-1}\Big[\, B_1\boldsymbol{S}_\zeta B_1^{\mathrm{T}} + \boldsymbol{B}_2\boldsymbol{S}_\zeta\boldsymbol{B}_2^{\mathrm{T}} + \mathrm{e}^{i\omega(t_1-t_2)}\,\boldsymbol{B}_2\boldsymbol{S}_\zeta\boldsymbol{B}_1^{\mathrm{T}} + \mathrm{e}^{i\omega(t_2-t_1)}\,\boldsymbol{B}_1\boldsymbol{S}_\zeta\boldsymbol{B}_2^{\mathrm{T}}\Big](-i\omega\boldsymbol{E} - \boldsymbol{A})^{-\mathrm{T}}. \tag{29.100}$$

In the frequency domain, the time delay $(t_2 - t_1)$ results in additional terms, weighted by the exponential function.

29.3.3 Covariance Analysis

In contrast to spectral analysis, the covariance analysis directly yields the variances that are required for the assessment of ride comfort and safety. The covariance matrix of the entire vehicle system follows from an algebraic equation, the so-called Lyapunov matrix equation, and no integrations are required. An essential prerequisite of the covariance analysis is a white noise input process. This can always be provided by modeling the random vehicle excitation by means of a shape filter.

The Lyapunov matrix equation corresponding to the state Equation 29.92 of the entire vehicle system reads

$$\boldsymbol{A}\boldsymbol{P}_x + \boldsymbol{P}_x\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{Q}_W\boldsymbol{B}^{\mathrm{T}} = 0, \tag{29.101}$$

where

$\boldsymbol{P}_x = E\,\{\boldsymbol{x}\,\boldsymbol{x}^{\mathrm{T}}\}$ denotes the symmetric $n \times n$-covariance matrix

$\boldsymbol{Q}_W$ is the $r \times r$-intensity matrix of the white noise input

An extensive derivation of Equation 29.101 can be found in [20]. Stable numerical procedures to solve the Lyapunov matrix equation are given in [21,22]. As a result, one gets the variances of all state variables of the system. However, the computation of the ride comfort requires the variance of the frequency weighted acceleration. Thus, the covariance analysis has to be applied to the extended system (29.94). The covariance matrix of the extended state vector $\tilde{\boldsymbol{x}}$ reads

$$\boldsymbol{P}_{\tilde{x}} = \begin{bmatrix} \boldsymbol{P}_x & \boldsymbol{P}_{x\bar{v}} \\ \boldsymbol{P}_{\bar{v}x} & \boldsymbol{P}_{\bar{v}} \end{bmatrix} \tag{29.102}$$

where $\boldsymbol{P}_{x\bar{v}} = \boldsymbol{P}_{\bar{v}x}^{\mathrm{T}}$ holds. The $s \times 1$-vector process $\bar{\boldsymbol{v}}(t)$ yields, according to (29.21), immediately the frequency weighted acceleration and its variance

$$\sigma_{\bar{a}}^2 = \alpha^2\,\bar{\boldsymbol{h}}^{\mathrm{T}}\boldsymbol{P}_{\bar{v}}\,\bar{\boldsymbol{h}}. \tag{29.103}$$

This is the relation corresponding to Equation 29.98. The investigation of multiaxle vehicles is also possible by the covariance analysis. The theory needed can be developed, as shown in [16]. In contrast to spectral analysis, the covariance analysis can also be applied to nonstationary and nonlinear problems without difficulties.

Example 29.5

Random vibrations of a single-axle vehicle.

A wheel guided in the vertical direction is excited by the road roughness and performs random vibration (Figure 29.15). The variance of the dynamic wheel load variations have to be investigated by applying covariance analysis. The road roughness is characterized by white velocity noise $\dot{\zeta}(t) \sim (0, q)$.

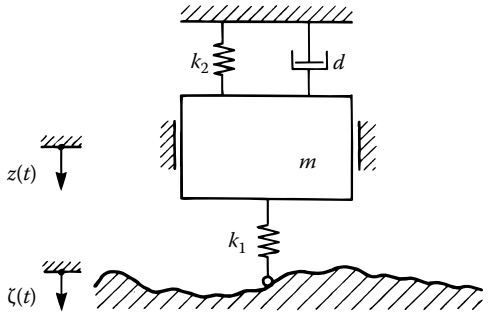


FIGURE 29.15 Single-axle vehicle model under random excitation.

Solution:

The equation of motion of the wheel reads

$$m\ddot{z} + d\dot{z} + (k_1 + k_2)z = k_1\zeta(t), \tag{29.104}$$

where $\zeta(t)$ is a scalar random process. The corresponding state equation has the form

$$\underbrace{\begin{bmatrix} \dot{z} \\ \ddot{z} \end{bmatrix}}_{\dot{\boldsymbol{x}}} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\kappa_{12} & -\delta \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} z \\ \dot{z} \end{bmatrix}}_{\boldsymbol{x}} + \underbrace{\begin{bmatrix} 0 \\ \kappa_1 \end{bmatrix}}_{\boldsymbol{B}} \zeta(t), \tag{29.105}$$

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{w}(t),$$

where the abbreviations $\delta = d/m$, $\kappa_{12} = (k_1 + k_2)/m$, $\kappa_1 = k_1/m$ have been used.

Since the random excitation is given by a white velocity noise process, it is recommendable to differentiate Equation 29.105 so that the process $\dot{\zeta}(t)$ appears directly as excitation. Then, the Lyapunov equation reads

$$\boldsymbol{A}\boldsymbol{P}_{\dot{x}} + \boldsymbol{P}_{\dot{x}}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{q}\boldsymbol{B}^{\mathrm{T}} = 0, \tag{29.106}$$

where the 2 × 2-covariance matrix of the first derivative $\dot{x}$ of the state vector has been introduced,

$$\boldsymbol{P}_{\dot{x}} = \begin{bmatrix} P_{11} & P_{12} \\ P_{12} & P_{22} \end{bmatrix} = \begin{bmatrix} P_{\dot{z}\dot{z}} & P_{\dot{z}\ddot{z}} \\ P_{\dot{z}\ddot{z}} & P_{\ddot{z}\ddot{z}} \end{bmatrix}. \tag{29.107}$$

The solution of Equation 29.106 yields the following matrices

$$\boldsymbol{A}\boldsymbol{P}_{\dot{x}} = \begin{bmatrix} P_{12} & P_{22} \\ -\kappa_{12}P_{12} - \delta P_{12} & -\kappa_{12}P_{12} - \delta P_{22} \end{bmatrix}, \tag{29.108}$$

$$\boldsymbol{B}\boldsymbol{q}\boldsymbol{B}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 0 & \kappa_1^2 q \end{bmatrix}. \tag{29.109}$$

Due to the symmetry of the covariance matrix from Equation 29.106, it follows altogether three linear equations for the unknowns

$$\begin{bmatrix} 0 & 2 & 0 \\ -\kappa_{12} & -\delta & 1 \\ 0 & -2\kappa_{12} & -2\delta \end{bmatrix} \begin{bmatrix} P_{11} \\ P_{12} \\ P_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \kappa_1^2 q \end{bmatrix} \tag{29.110}$$

with the solutions

$$P_{11} = \frac{k_1^2}{2d(k_1 + k_2)} q, \quad P_{12} = 0, \quad P_{22} = \frac{k_1^2}{2dm} q. \tag{29.111}$$

The weighting vector b of the dynamic wheel loads reads, according to Equation 29.104,

$$f_{Dyn} = k_1(z - \zeta) = \underbrace{[-d - m]}_{\boldsymbol{b}^T} \underbrace{\begin{bmatrix} \dot{z} \\ \ddot{z} \end{bmatrix}}_{\dot{x}}, \tag{29.112}$$

where $k_2 << k_1$ has been regarded. The variance σ_f^2 of the dynamic wheel load variation reads

$$\sigma_f^2 = \boldsymbol{b}^T \boldsymbol{P}_{\dot{x}} \boldsymbol{b} = \left[\frac{k_1 d}{2} + \frac{k_1^2 m}{2d} \right] q. \tag{29.113}$$

It can be seen that σ_f^2 becomes a minimum if the viscous damping factor $D = d \Big/ \left(2\sqrt{k_1 m} \right) = 0.5$ is chosen. In this case of optimal damping, it follows $\sigma_f^2 = k_1 \sqrt{k_1 m} q$. This result shows that small wheel masses m and small tire stiffnesses k_1 are beneficial for the dynamic wheel load variations. However, there are design limits for these parameters (see Chapter 8).

The assumption $k_2 << k_1$ is necessary in this example, since an idealized white velocity noise process has been considered. In the general case of colored noise, this assumption is not required.

Example 29.6

Random vibrations of a two-axle vehicle.

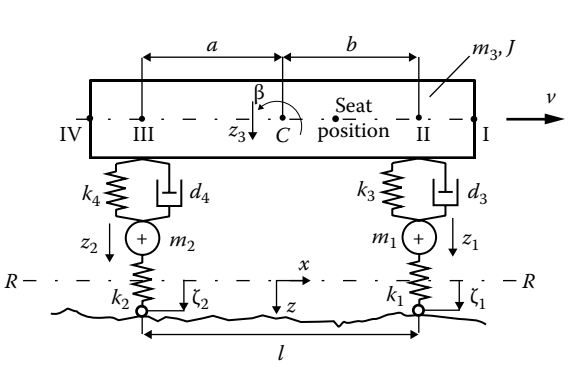
The covariance analysis is demonstrated by the example shown in Figure 29.16. A two-axle vehicle with four degrees of freedom is excited by colored noise $\zeta(t)$ according to (29.79) or white noise (29.81) for the derivated process $\dot{\zeta}(t)$.

Small motions are assumed and the time lag $t_2 = l_2/v$ between the front and rear contact has to be taken into account.

Solution:

The generalized coordinates are summarized in the position vector $\boldsymbol{y}(t)$, $\boldsymbol{y}(t) = [y_1(t), y_2(t), y_3(t), \Phi(t)]^T$. Then, the equation of motion reads

$$\boldsymbol{M}\ddot{\boldsymbol{y}}(t) + \boldsymbol{D}\dot{\boldsymbol{y}}(t) + \boldsymbol{K}\boldsymbol{y}(t) = \boldsymbol{s}_1\zeta_1(t) + \boldsymbol{s}_2\zeta_2(t), \quad \zeta_1(t) = \zeta(t), \quad \zeta_2(t) = \zeta(t - t_2), \tag{29.114}$$



$m_1 = 100$ kg
$m_2 = 54$ kg
$m_3 = 1{,}247$ kg
$J = 1{,}945$ kgm^2
$k_1 = k_2 = 34{,}3350$ N/m
$k_3 = 47{,}382$ N/m
$k_4 = 51{,}208$ N/m
$d_3 = 3{,}198$ Ns/m
$d_4 = 3{,}434$ N/m
$a = 1.2$ m
$b = 1.3$ m
$l = 2.5$ m

FIGURE 29.16 Two-axle vehicle model under random excitation and system parameters.

where $\boldsymbol{M}$, $\boldsymbol{D}$, and $\boldsymbol{K}$ characterize inertia, damping, and stiffness matrix, respectively, and $\boldsymbol{s}_1$, $\boldsymbol{s}_2$ are the excitation input vectors, $\boldsymbol{M} = \mathbf{diag}[m_1, m_2, m_3, J]$, $\boldsymbol{s}_1 = [k_1 0\ 0\ 0]^T$, $\boldsymbol{s}_2 = [0\ k_2 0\ 0]^T$,

$$\boldsymbol{D} = \begin{bmatrix} d_3 & 0 & -d_3 & bd_3 \\ 0 & d_4 & -d_4 & -ad_4 \\ -d_3 & -d_4 & d_3 + d_4 & -bd_3 + ad_4 \\ bd_3 & -ad_4 & -bd_3 + ad_4 & b^2 d_3 + a^2 d_4 \end{bmatrix}, \tag{29.115}$$

$$\boldsymbol{K} = \begin{bmatrix} k_1 + k_3 & 0 & -k_3 & bk_3 \\ 0 & k_2 + k_4 & -k_4 & -ak_4 \\ -k_3 & -k_4 & k_3 + k_4 & -bk_3 + ak_4 \\ bk_3 & -ak_4 & -bk_3 + ak_4 & b^2 k_3 + a^2 k_4 \end{bmatrix}. \tag{29.116}$$

From (29.114), the vehicle state space representation can be obtained,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}_1\zeta_1(t) + \boldsymbol{b}_2\zeta_2(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\xi, \tag{29.117}$$

$$\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{y}(t) \\ \dot{\boldsymbol{y}}(t) \end{bmatrix}, \quad \xi(t) = \begin{bmatrix} \zeta_1(t) \\ \zeta_2(t) \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{E} \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{D} \end{bmatrix}, \quad \boldsymbol{b}_1 = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{M}^{-1}\boldsymbol{s}_2 \end{bmatrix}$$

$$\boldsymbol{B} = [\boldsymbol{b}_1 \quad \boldsymbol{b}_2].$$

In order to use white noise vehicle excitation, $\dot{\zeta}(t) \equiv w(t)$, Equation 29.116 has to be differentiated,

$$\left[\dot{\boldsymbol{x}}(t)\right]^{\bullet} = \boldsymbol{A}\left[\dot{\boldsymbol{x}}(t)\right] + \boldsymbol{B}\boldsymbol{w}(t). \tag{29.118}$$

Here, the autocorrelation matrix of the white noise vector process w reads

$$\boldsymbol{R}_W(\tau) = E\left\{\boldsymbol{w}(t)\boldsymbol{w}^T(t-\tau)\right\} = q_\zeta \begin{bmatrix} \delta(\tau) & \delta(\tau + t_2) \\ \delta(\tau - t_2) & \delta(\tau) \end{bmatrix}. \tag{29.119}$$

Thus, the covariance matrix $\boldsymbol{P}_{\dot{x}}$ in the steady state follows from the algebraic Lyapunov equation

$$\boldsymbol{A}\boldsymbol{P}_{\dot{x}} + \boldsymbol{P}_{\dot{x}}\boldsymbol{A}^T + \boldsymbol{Q} = \boldsymbol{0}, \quad \mathrm{c}\,\boldsymbol{Q} = q_{\dot{\zeta}}\left[\boldsymbol{b}_1\boldsymbol{b}_1^T + \boldsymbol{b}_2\boldsymbol{b}_2^T + e^{At_2}\boldsymbol{b}_1\boldsymbol{b}_2^T + \boldsymbol{b}_2\boldsymbol{b}_1^T e^{\boldsymbol{A}^T t_2}\right]. \tag{29.120}$$

From $\boldsymbol{P}_{\dot{x}}$, the variances σ_a^2 of the car body acceleration $\ddot{y}_p(t)$ at any point P with coordinates $y_p(t) = \boldsymbol{a}^T\boldsymbol{y}(t)$ can immediately be calculated by

$$\sigma_a^2 = [\boldsymbol{0}^T\boldsymbol{a}^T]\ \boldsymbol{P}_{\dot{x}}\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{a} \end{bmatrix}. \tag{29.121}$$

A similar expression holds for the variances σ_F^2 of the wheel loads $F_i = k_i(\zeta_i - y_i)$, $i = 1, 2$, but here Equation 29.114 has to be used to eliminate the displacements. Figure 29.17 shows some numerical results, where a smooth road ($q_\zeta = 3{,}14{\cdot}10^{-6} \cdot \mathrm{vm}^2/s$ – solid line) and a rough road excitation ($q_\zeta = 24.7{\cdot}10^{-6} \cdot \mathrm{vm}^2/s$ – dashedline) is assumed. In Figure 29.17a), the standard deviation σ_F/F_o of the load

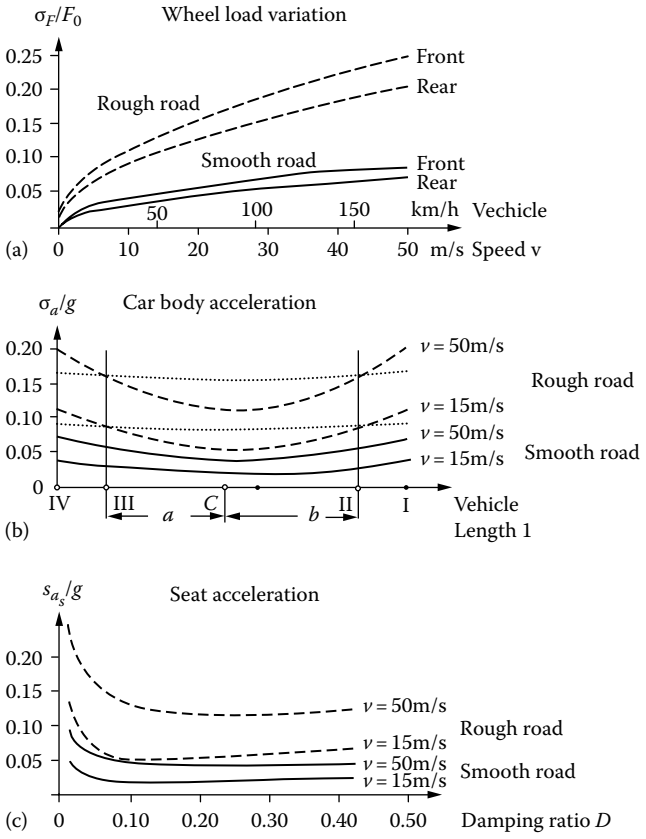


FIGURE 29.17 Random response variances, (a) wheel load variation depending on vehicle speed, (b) car body acceleration over vehicle length, (c) seat acceleration versus damping (— — rough road — — smooth road).

variation of the front and rear wheels is plotted against speed v, where F_o denotes the static weight. It increases with $\sqrt{v}$ since the excitation intensity is proportional to v. Even in the worst case (rough road, front wheel, maximum speed), the safety margin R according to Equation 29.28 takes the value $R = 1-\sigma_F/F_o \approx 0.75$, which indicates a safe ride. In Figure 29.17b, the standard deviation σ_a/g of the car body acceleration versus vehicle length is given, which clearly shows a minimum for the center of mass position at different speeds. Neglecting the time lag between the front and rear excitation (dotted line) can lead to incorrect results, especially near the seat position. The covariance analysis has an important application in optimization problems also. Figure 29.17c shows a simple example, where the influence of the nondimensional damping ratio $D = D_i \equiv d_i / \left(2\sqrt{k_i \bar{m}_i}\right)$, $i = 1,2,(\bar{m}_1 = m_3 a/l_1$, $\bar{m}_2 = m_3 b/l_1)$ on the standard deviation of the seat acceleration, σ_{a_s}/g, is studied. Best results are obtained in the range $0.1 \le D \le 0.25$. However, tradeoffs with the wheel load variation call for higher damping. For ride comfort evaluation, the perception measure K has to be calculated. A rough approximate estimation using the thumb rule $a_{max} \le 0.5$ m/s shows that the worst case (rough road, maximum speed, and maximum damping) certainly results in bad ride comfort. For further information on optimization methods of vehicle suspensions, see Chapter 8.

REFERENCES

1. Müller, P. C.: *Stabilität und Matrizen*. Berlin, Germany: Springer, 1977.
2. ISO International Standard 2631: *Guide for the Evaluation for the Human Exposure to Whole-Body Vibrations*. Berlin, Germany: Beuth, 1974. Also ISO 2631: Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole-Body Vibrations. Berlin, Germany: Beuth, 1997–2004.

3. VDI–Richtlinie 2057: *Beurteilung der Einwirkung mechanischer Schwingungen auf den Menschen*. Duesseldorf, Germany: VDI-Verlag, 1975–1979. Also VDI 2057: Einwirkung mechanischer Schwingungen auf den Menschen-Ganzkoerper-Schwingungen. Duesseldorf, Germany: VDI-Verlag, 2002.
4. Popp, K.: Schiehlen, W.: Ground Vehicle Dynamics. Berlin, Germany: Springer, 2010.
5. Wong, J. Y.: *Theory of Ground Vehicles*. New York: Springer, 2001.
6. Mitschke, M.: *Dynamik der Kraftfahrzeuge*. Berlin, Germany: Springer, 1972.
7. Braun, H.: *Untersuchungen von Fahrbahnunebenheiten und Anwendung der Ergebnisse*. Braunschweig, Germany: Dr.-Ing. Diss., 1969.
8. Voy, C.: *Die Simulation vertikaler Fahrzeugschwingungen*. Fort. Ber. VDI Reihe12 Nr. 30. Düsseldorf, Germany: VDI-Verlag, 1977.
9. Bormann, V.: Messungen von Fahrbahnunebenheiten paralleler Fahrspuren und Anwendung der Ergebnisse. *Vehicle Syst. Dyn*. 7 (1978), 65–81.
10. Crandall, S. H.; Mark, W. D.: *Random Vibration in Mechanical Systems*. New York, London, U.K.: Academic Press, 1963.
11. Newland, D. E.: *Random Vibrations and Spectral Analysis*. London, U.K., New York: Longman, 1975.
12. Heinrich, W., Hennig, K.: *Zufallsschwingungen mechanischer Systeme*. Braunschweig, Germany: Vieweg, 1978.
13. Dodds, C. J., Robson, J. D.: The description of road surface roughness. *J. Sound Vibration* 31 (1973), 175–183.
14. Rill, G.: *Instationäre Fahrzeugschwingungen bei stochastischer Erregung*. Stuttgart, Germany: Dr.-Ing. Diss., 1983.
15. Czerny, L.: *Analyse instationärer Zufallsschwingungen*. Fort. Ber. VDI Reihe 11 Nr. 99. Düsseldorf, Germany: VDI-Verlag, 1987.
16. Müller, P. C., Popp, K., Schiehlen, W. O.: Berechnungsverfahren für stochastische Fahrzeugschwingungen. *Ing. Arch*. 49 (1980), 235–254.
17. Müller, P. C., Popp, K.: Kovarianzanalyse von linearen Zufallsschwingungen mit zeitlich verschobenen Erregerprozessen. *Z. angew. Math. Mech*. 59 (1979), T144–T146.
18. Rill, G.: *Auswahl eines geeigneten Integrationsverfahrens für nichtlineare Bewegungsgleichungen bei Erregung durch beliebige Zeitfunktionen*. Stuttgart, Germany: Forschungsbericht FB-4. Inst. B f. Mechanik, 1981. Also Rill, G.; Schiehlen, W.: Performance assessment of time integration methods for vehicle dynamics simulations In: Arczewski, K.; Frazek, J.; Wojtyra, M. (eds.): Multibody Dynamics 2009, ECCOMAS Thematic Conference. Warsaw, Poland: Warsaw University of Technology, 2009.
19. Shampine, L. F., Gordon, M. K.: *Computer Solution of Ordinary Differential Equations*. San Francisco, CA: Freeman, 1984: Braunschweig, Germany: Vieweg, 1984.
20. Müller, P. C.; Schiehlen, W.: *Lineare Schwingungen*. Wiesbaden, Germany: Akademische Verlagsgesellschaft, 1976 und Stuttgart, Germany: Inst. B. f. Mechanik (ISBN 3-927618-13-6), 1991.
21. Smith, R. A.: Matrix equation XA + BX = C. *SIAM J. Appl. Math*. 16 (1968), 198–201.
22. Kreisselmeier, G.: A solution of the bilinear matrix equation AY + YB = -Q. *SIAM J. Appl. Math*. 23 (1972), 334–338.

30 Control of Horizontal Vehicle Motion

Anton van Zanten

CONTENTS

30.1 Overview of Control Systems of Cars 1094
30.1.1 Controllability of Cars 1094
30.1.2 Basic Tire Properties 1095
30.1.3 The Antilock Brake System 1100
30.1.4 Traction Slip Control System 1108
30.1.5 Electronic Stability Program 1111
30.1.5.1 System Concept 1113
30.1.5.2 Performance Requirements 1118
30.1.5.3 Vehicle Dynamics Controller 1120
30.1.5.4 Brake Slip Controller 1129
30.1.5.5 Traction Slip Controller 1134
30.1.5.6 Results 1140
30.2 Special Functions of ESP 1143
30.2.1 Off-Road Vehicles 1143
30.2.1.1 ABS Braking Off Road 1143
30.2.1.2 Hill Descent Control 1145
30.2.2 Rollover Mitigation 1145
30.2.3 Trailer Oscillation Mitigation 1145
30.2.4 Electronic Brake Force Distribution 1146
30.2.5 Brake Assistant 1147
30.3 ESP Safety Concept 1149
30.3.1 Requirements of the Safety System 1150
30.3.2 Avoidance of Failures 1151
30.3.3 Self-Test, Self-Monitoring, and Tests of Components 1152
30.3.4 Failure Detection Logic 1153
30.3.4.1 Model-Based Sensor Monitoring 1154
30.3.4.2 Measures for Unrecognized Failures 1157
30.3.5 System Behavior after Detection of Failure 1157
30.3.6 Fallback Functions 1157
30.4 Components 1158
30.4.1 Sensors 1158
30.4.1.1 Yaw Rate Sensor 1158
30.4.1.2 Steering Wheel Angle Sensor 1161
30.4.1.3 Pressure Sensor 1162

30.4.2 Actuators....1162
30.4.2.1 Hydraulic Unit....1162
30.4.2.2 Precharge Pump eVLP....1164
30.4.3 Electronic Control Unit....1164
30.5 Outlook....1165
Nomenclature....1168
Acronyms....1172
References....1174

30.1 OVERVIEW OF CONTROL SYSTEMS OF CARS

In some driving situations, the driver may lose control over his car. For instance, such a situation may arise if the driver applies the brake pedal too hard and the wheels stop rotating and lock up. If the front wheels are locked, the driver will not be able to steer the car and the car will drift, while if the rear wheels are locked, the driver will not be able to keep the car from swerving. A similar situation may arise if the driver applies the accelerator pedal too much: the wheels spin with the same effect on the car as if the wheels were locked. However, even if neither the accelerator pedal nor the brake pedal is applied too much, the driver may still lose control over his car, for example, because of violent steering maneuvers.

In those situations, control systems help the driver to maintain control over his car. These control systems are the antilock brake system (ABS), which keeps the wheels from locking; the traction slip control system (TCS or ASR), which keeps the wheels from spinning; and the electronic stability program (ESP), which keeps the car from swerving and drifting. In the following sections, ABS and TCS will be explained as stand-alone systems. ESP includes the functions ABS and TCS which will be explained as parts of the hierarchical structure of ESP. For an overview of these systems, see [1–3].

30.1.1 Controllability of Cars

It is well known that if the wheels of a car lock or spin, then the resultant forces and the yaw moment on the car are independent of the slip angles of the tires. Thus, the loss of controllability of the car can be explained. However, it is not immediately clear why a car can become uncontrollable even if the wheels do neither lock nor spin (Figure 30.1).



FIGURE 30.1 Swerving car on dry asphalt, initial velocity 100 km/h, step steering wheel angle input 90°.

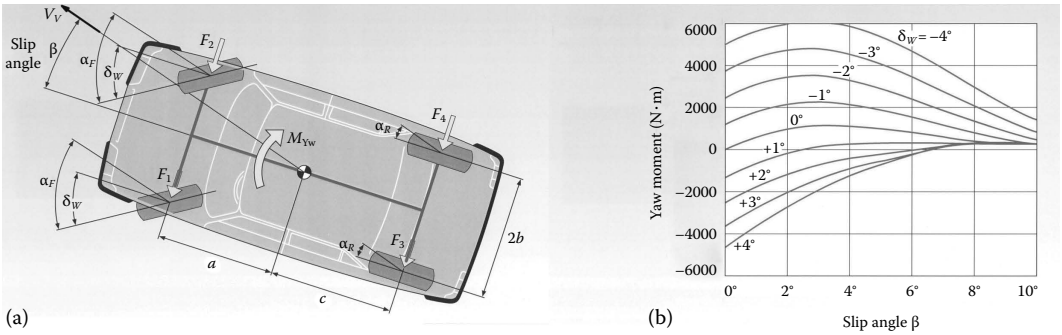


FIGURE 30.2 The beta method: (a) Top view of the car, (b) yaw moment on the car in dependence of the slip angle of the car for various steering wheel angles.

In order to explain this, Honda has developed a method [4] called the beta method, which shows the influence of the slip angle of the car on the yaw moment (Figure 30.2). For small slip angles, the yaw moment changes due to changes in the steering angle is much larger than for large slip angles. If for some reasons, the slip angle becomes large, then the influence of changing the steering angle on the yaw moment is small and may be so small that the slip angle cannot be reduced by steering and car swerve can hardly be avoided. This explains why average drivers that usually do not steer more than 90° at the steering wheel in either direction lose control if the slip angle of the car is large (larger than 6°).

Another explanation of why a car may become uncontrollable is given by Toyota (Figure 30.3) [5]. The figures show the trajectories of the car motion for combinations of slip angle and slip angle velocity at different steering angles. In the stable area, combinations of slip angle and slip angle velocity lead to a steady state of constant slip angle and zero slip angle velocity. If the steering angle is zero, then the steady state is a zero slip angle and a zero slip angle velocity. This means that eventually the car will move straight ahead, independent of the initial state (Figure 30.3a). In the unstable area, the steady state is not achieved and the car will not drive straight ahead, which, however, is the intention of the driver. If the steering angle is not zero, then the stability area is shrunk (Figure 30.3b), and if the steering angle is large enough (Figure 30.3c), there is even no stable state of the vehicle in motion.

Controlling the car in such situations may be extremely difficult for the driver. Here, ESP helps him with his task of keeping the car under control. Furthermore, the results show that even if ABS and TCS keep the wheels from locking and from spinning, respectively, it may still be difficult to control the car. Special control logic is used in these systems to improve the controllability of the car while keeping the wheels from locking and spinning, respectively. Furthermore, the systems must be tuned to each car to achieve optimum performance with regard to the car behavior. This shows the necessity of ESP to improve vehicle controllability for ABS and TCS situations also.

30.1.2 Basic Tire Properties

If a wheel is not steered and neither braked nor driven, then its rotational velocity, called the free rolling rotational velocity ω_{R0}, can be computed from the vehicle velocity v_v: $\omega_{R0} = v_v/R$ (Figure 30.4). The wheel radius R is computed by dividing the distance the wheel has traveled during one complete revolution by 2π. If the wheel is braked by a brake torque M_{BR}, then the rotational velocity ω_R will be smaller than the free rolling rotational velocity ω_{R0}. In the description of the control systems, the wheel velocity is used instead of the rotational velocity. The wheel velocity is defined as the product of the rotational wheel velocity and the wheel radius R. The free rolling wheel velocity is then $v_{WhlFre} = \omega_{R0} R$, while the wheel velocity is $v_{Whl} = \omega_R R$.

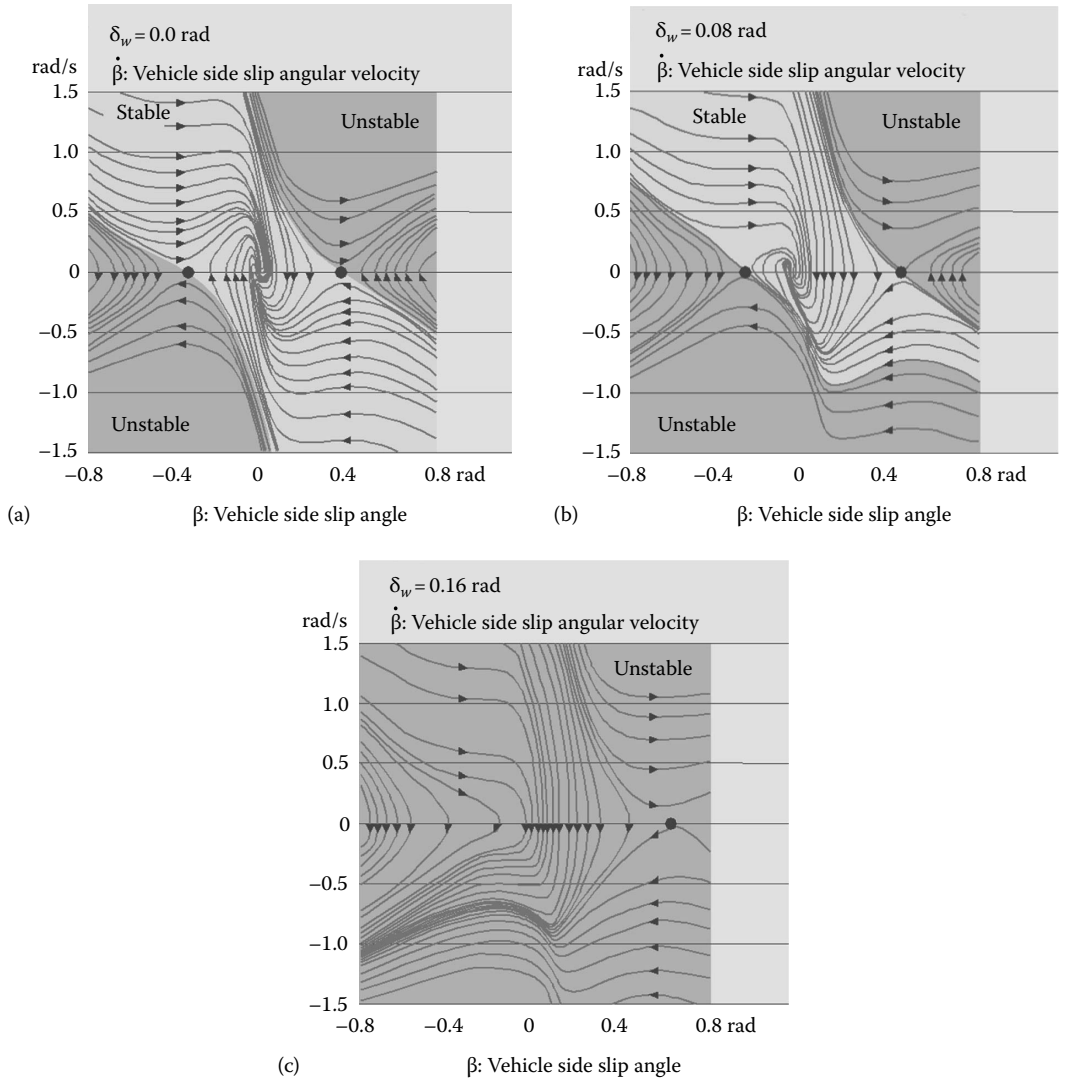


FIGURE 30.3 Phase plane plots of the car motion with stable and unstable regions depending on the steering wheel angle. (a) Steering wheel angle 0 rad, (b) Steering wheel angle 0.08 rad, (c) Steering wheel angle 0.16 rad.

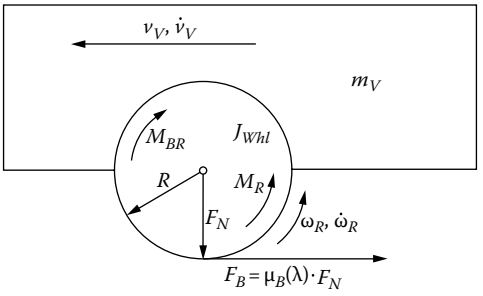


FIGURE 30.4 Simple physical model of a wheel attached to the vehicle body.

The difference between the free rolling and the braked wheel velocity is called the slip velocity. The slip velocity divided by the free rolling wheel velocity is called the wheel or tire slip λ. It is a dimensionless quantity whose value is one if the wheel is locked. Often, the slip is also expressed in percent, and the slip is 100% if the wheel locks.

$$\lambda = \frac{v_{WhlFre} - v_{Whl}}{v_{WhlFre}} \tag{30.1}$$

If a brake torque M_{BR} acts on the wheel, then a brake force F_B between the tire and the road results. If the normal force on the tire is F_N, then the brake coefficient of friction is $\mu_B = F_B/F_N$. The road torque on the wheel M_R is defined as the product of the brake force F_B and the wheel radius R: $M_R = \mu_B \cdot F_N \cdot R$. Between the brake slip and the brake coefficient of friction, a nonlinear relationship exists, called the μ-slip curve. Figure 30.5 shows a typical μ-slip curve, while Figure 30.6 shows some slip curves for various road surfaces. The curves usually exhibit a maximum, but for loose road surfaces

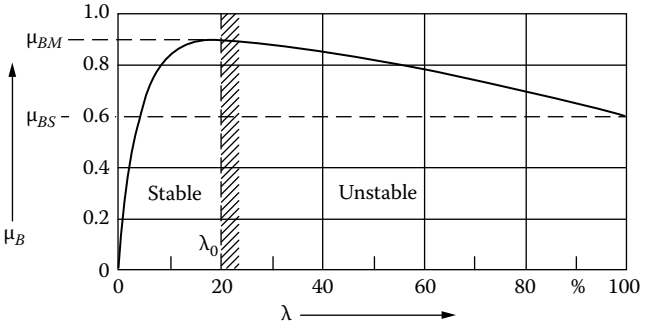


FIGURE 30.5 Typical μ-slip curve with μ_{BM} as the maximum value of the coefficient of friction and μ_{BS} as the value of the coefficient of friction of the skidding tire.

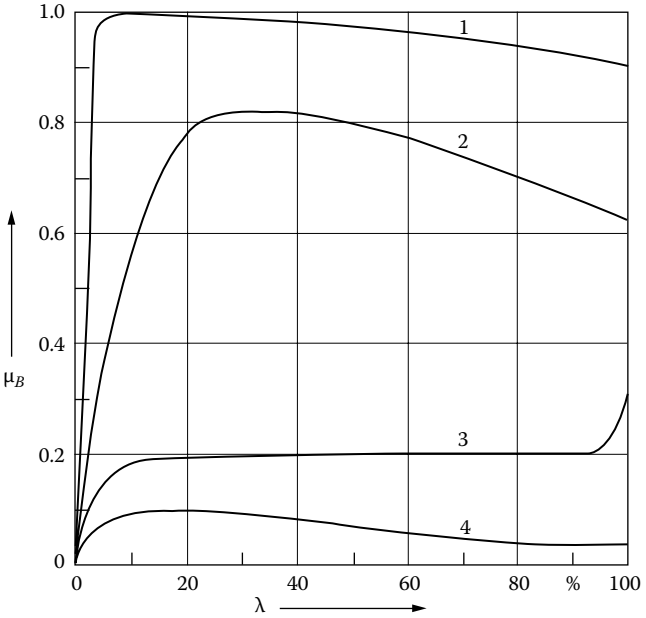


FIGURE 30.6 Typical μ-slip curves for some road surfaces: (1) dry asphalt, (2) wet asphalt, (3) packed snow, and (4) polished ice.

like snow and sand, the curve may not exhibit a maximum. The slip value at the maximum of the slip curve is often called λ_0. For values of the tire slip larger than λ_0, the μ-slip curve is unstable, since there is no stable equilibrium between the brake torque and the road torque: the tire slip will not be stable for a constant brake torque but will increase until the wheel locks.

If a free rolling wheel, whose longitudinal velocity is v_x, is pushed sideways, then the wheel will move also in the lateral direction and the resultant velocity of the wheel center is v_{VWhl} (Figure 30.7). The angle between the resulting wheel velocity and the wheel plane is called the slip angle α. Because the wheel is pushed sideways, the road acts with a force F_S in the opposite lateral direction on the wheel. The side force depends on the slip angle and on the normal force on the tire. Figure 30.8 shows the side force to be dependent on the slip angle for some wheel loads. Note that the maximum side force is not proportional to the normal force. The lateral coefficient of friction is the lateral force divided by the normal force on the wheel, $\mu_S = F_S/F_N$. The nonlinear relationship between the slip angle and the lateral coefficient of friction is called the μ-slip angle curve, which looks similar to that between the slip and the longitudinal coefficient of friction.

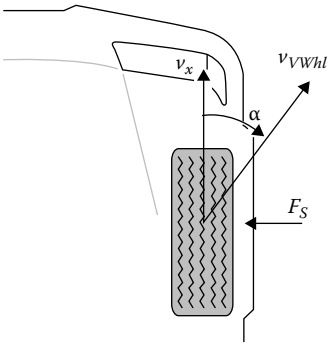


FIGURE 30.7 Definition of the slip angle and the side force on a wheel.

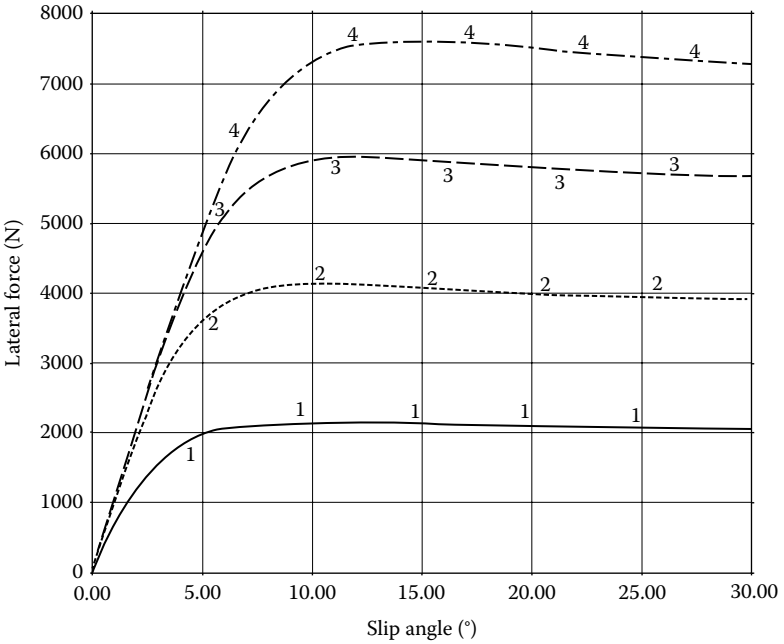


FIGURE 30.8 Lateral force on the wheel depending on the slip angle of the wheel for some wheel loads: (1) 2000 N, (2) 4000 N, (3) 6000 N, and (4) 8000 N.

The lateral coefficient of friction depends also on the tire slip: in general, if the tire slip is increased, then the lateral coefficient of friction drops if the slip angle is kept fixed (Figure 30.9). Furthermore, the μ-slip curve has different shapes for different slip angles. In general, μ_B decreases for a fixed slip value and λ_0 increases if the slip angle increases.

Often the relationship between the longitudinal and the lateral force on the wheel is shown as a map, in which curves of constant slip angle values and variable slip values are plotted (Figure 30.10). If the wheel locks, then the resultant force between the tire and the road has a direction opposite to

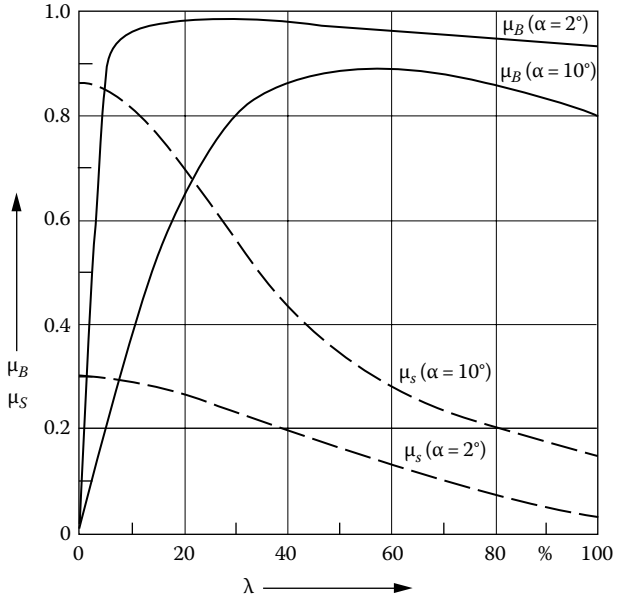


FIGURE 30.9 μ-slip curves for some fixed slip angles and the side force coefficient of friction for some fixed slip angles depending on the slip.

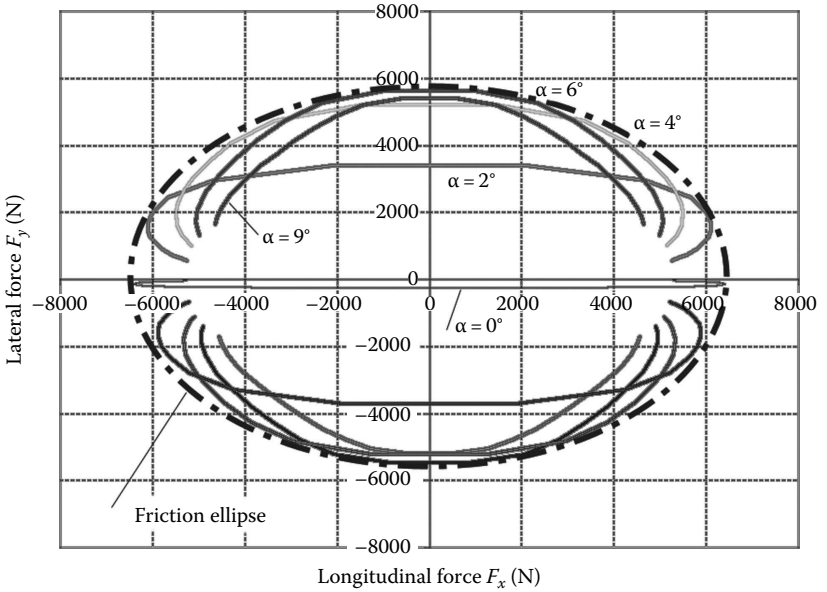


FIGURE 30.10 Map of the lateral and longitudinal forces for some fixed slip angles and the friction ellipse as the envelope of the map.

the velocity of the wheel: at those points in the map is $\tan \alpha = F_S/F_B$. For the control systems, the envelope of the map is approximated by an ellipse, which is called the friction ellipse.

30.1.3 The Antilock Brake System

The idea to keep the wheels from locking is very old and dates back to the beginning of the twentieth century. The first modern electronically controlled ABS was installed in a Mercedes car in 1978. An excellent description of the physics and the idea behind this system can be found in [6].

The first requirement of ABS is to keep the wheels from locking. However, the requirements can be formulated more specifically. The most important requirements are as follows:

1. The control of the brakes should secure a good controllability of the car by the driver on all roads, even if the road on one side of the car is more slippery than on the other side of the car (called split-μ braking).
2. In case of hydroplaning, stability and straight-line driving should be secured.
3. The coefficient of friction between the tires and the road should be maximally exploited during full braking.
4. The control function must be given at all car velocities down to zero.
5. The control performance must be tolerant of wear of the car and of the system, and also with respect to tire types and tire wear.
6. In the temperature range from −20°C to 120°C, the system must show full performance, while at lower temperatures the performance must not suddenly vanish.
7. Priorities during high-speed straight-line braking:
   a. Stopping distance
   b. Car stability
   c. Comfort (noise and brake pedal pulsation)
8. Priorities during high-speed handling:
   a. Car stability
   b. Stopping distance
   c. Comfort (noise and brake pedal pulsation)
9. Priorities during μ-split braking:
   a. Stopping distance
   b. Car stability
   c. Car steer ability (steering wheel rotational speed <90°/s)
   d. Car lateral displacement (<0.5 m relative to the middle of the track)
   e. Comfort (noise and brake pedal pulsation)
10. The system should be constantly monitored for proper performance. If a failure in a system component is recognized, which may reduce the system performance, then the system must switch off itself. The driver must be warned by a warning light on the display of the dashboard. If the system is switched off, then the base brake system must be fully functional.

In order to meet these requirements, the control system must limit the slip of the wheels such that the lateral stability margin at the front and at the rear axles is large enough for the driver to keep his car under control while at the same time minimizing the stopping distance.

For cost reasons, the car velocity is not measured; hence, the slip can neither be directly determined nor controlled. Instead, the stand-alone ABS is primarily designed as an optimizer of the brake force in the following way. If the slip is larger than the slip at the maximum brake force, λ_0, then the slip is unstable and the wheel will tend to lock, which is shown by a large deceleration of the wheel. And vice versa, it is assumed that if the deceleration of the wheel is large, then the slip is larger than λ_0 and the brake force is smaller than that at the maximum. In this situation, the brake pressure must be reduced in order to accelerate the wheel, increase its velocity, and thus reduce its

slip and increase its brake force. Furthermore, it is assumed that after the acceleration of the wheel has dropped to a small value, then the slip is smaller than λ_0 and the brake force is smaller than its maximum value. In this situation, the brake pressure must be increased in order to increase the slip and thus increase the brake force. The brake pressure control at the wheel is thus governed by the acceleration and deceleration of the wheel: ABS primarily controls the acceleration and deceleration of the wheel. At the slip value λ_0, the lateral stability margins of the tires already improve the controllability of the vehicle during full braking.

ABS reduces the brake pressures at the wheels from that induced by the driver such that the wheels do not lock. If the driver does not apply the brake pedal hard enough and the wheel does not tend to lock, then the control does not start. Therefore, the brake pressure at the wheels cannot be larger than that induced by the driver: ABS cannot actively increase the pressure at the wheels.

The simplified principle of the pressure control of ABS is shown in Figure 30.11. In phase 1, the brake pressure at the wheel p_{Whl} at the beginning of a braking maneuver is shown. Because of the brake torque, the wheel decelerates and so does the vehicle. If the wheel acceleration a_{Whl} drops below the threshold $-a$, then the brake pressure increase is interrupted. It is kept constant and not immediately reduced since horizontal axle motions can feign a wheel instability. This may happen in particular at the beginning of a braking maneuver, when the compliance of the wheel carrier plays a role. Only after the wheel velocity has also dropped below the curve λ_1, the pressure is reduced. The curve λ_1 lies at a certain threshold value below the curve of the reference velocity v_{Ref}.

It is assumed that during the wheel deceleration, if a_{Whl} reaches the value $-a$, the coefficient of friction has reached its maximum value. Since v_{Ref} should approximate the wheel velocity at which the coefficient of friction is maximal, the reference velocity is extrapolated from this wheel velocity. The slope of the extrapolation is variable, starts with the value of −0.3 g (which corresponds approximately to the coefficient of friction of packed snow), and is large during each pressure increase. Furthermore, the v_{Ref} is derived for each diagonal wheel pair, for example, for the front-left and the rear-right wheel; and at the rear axle, the reference velocities are tied together by some logic. It is important that the average slope of v_{Ref} is not too steep, while otherwise, the wheels may lock before the car has come to a stop.

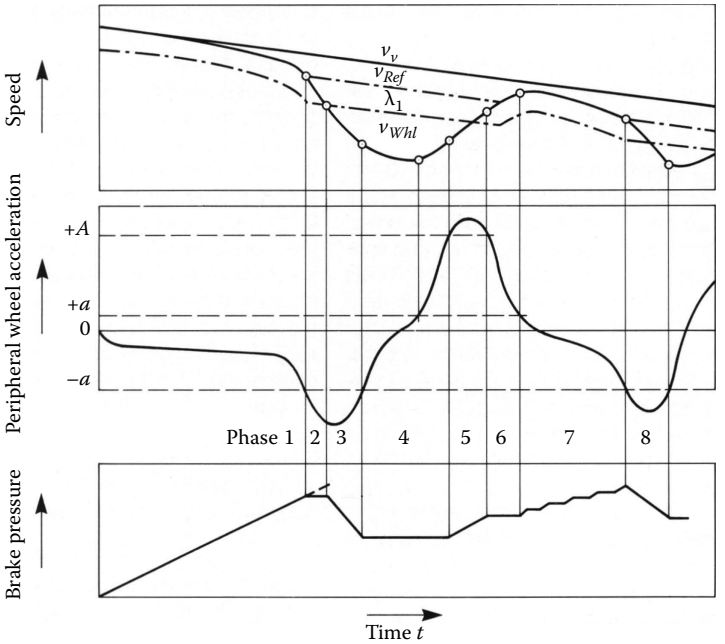


FIGURE 30.11 Typical initial part of the time plot of a stand-alone ABS braking control maneuver.

If the wheel deceleration diminishes and a_{Whl} increases above the threshold $-a$, the pressure reduction stops and the pressure is kept constant. During this phase (4), the wheel acceleration and velocity are observed. If the wheel velocity does not increase fast enough to become larger than the velocity λ_1, then the pressure is reduced in a stepwise manner. In phase 5, the acceleration of the wheel is so large that it exceeds the threshold $+A$. In this situation, it is assumed that the difference between the brake torque M_{BR} and the road torque M_R is very large. Therefore, the brake pressure is increased during this entire phase. After this phase 5, the brake pressure is kept constant and the wheel acceleration further observed.

If the wheel decelerates again, then a stable slip has established. Shortly before this happens, as the wheel acceleration drops below the threshold $+a$, the brake pressure is increased again—now in a stepwise manner. The first step is variable, while the following ones are fixed but increasing in magnitude. The value of the first step is chosen according to a learning scheme in order to get close to the maximum brake force. The idea behind following the stepwise manner, and thus a slow increase in the brake pressure, is to increase the slip level only slowly, such that the brake force stays at its maximum value for quite sometime. The pressure is increased stepwise until the wheel acceleration reaches the threshold $-a$ again. Now, the brake pressure is reduced immediately since it is assumed that the compliance of the wheel carrier has no influence on the wheel motion anymore. The control cycle is then repeated. Because of the logical nature of the control scheme with its many thresholds, it is often referred to as the ABS logic.

Since the wheel acceleration and deceleration depend only on the difference between the brake and the road torque and not on their magnitudes, the coefficient of friction of the road has little influence on the control performance. Therefore, ABS is adaptive to the road friction coefficient. The same is true for variations in the coefficient of friction of the brakes, for example, due to brake fading. The mean value of the slip can be modified by modifying the threshold values and by modifying the pressure steps. In general, for improved vehicle controllability, the modifications are chosen such that an average slip results, which is smaller than the slip value at the maximum of the μ-slip curve.

The control shown earlier aims at almost maximizing the brake force individually at each wheel. This type of control is called individual control. In some cases, individual control is not sufficient for the driver to keep his car under control. In particular, on split-μ roads (these are roads where on the one side of the car the coefficient of friction is smaller than on the other side of the car), special control features are required. These are "Select Low" control at the rear axle and "Yaw Moment Gradient Attenuation" (GMA) at the front axle. These special features will be discussed for a split-μ braking maneuver.

The braking maneuver on a split-μ road surface is shown in Figure 30.12. On the left side of the car, the coefficient of friction of the road surface is high ($\mu_B = 0.8$), while on the right side, it is very low ($\mu_B = 0.1$). For individual control, the brake forces on the left will be much higher than those on the right side. The result will be a large yaw moment on the car. The driver must balance this yaw moment by fast and heavy countersteering. In order to reduce the yaw moment and increase the stability margin at the rear axle of the car, the pressure on the left side of the rear axle is made equal to that on the right side (this is called the select low principle). Then the slip of the left rear wheel is very small and the side force reduction due to slip is also very small. Thus it is easier for the driver to control his car. This principle cannot be used at the front axle because then the higher coefficient of friction on the left side of the car would not be exploited at all and the stopping distance would be correspondingly long. Instead, at the front axle, a yaw moment is allowed; however, it is not allowed to increase very fast. This can be realized in several ways. One way is to start a stepwise brake pressure increase at the left wheel as soon as at the right wheel the deceleration threshold is reached (curve 3, Figure 30.13). The required steering effort is smaller (curve 7, Figure 30.13) than that for individual control (curves 2 and 6, Figure 30.13). Another way is shown in Figure 30.13 by curve 4. Because of this reduction in the gradient of the yaw moment, the driver has sufficient time to countersteer and keep his car stable on track.

If the car has a negative scrub radius, the countersteering effort of the driver is also reduced, since the wheels will automatically countersteer. However, both the GMA and the negative scrub radius

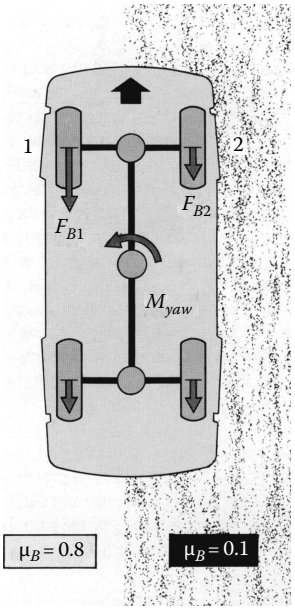


FIGURE 30.12 Brake forces F_{B1} and F_{B2} and yaw moment M_{yaw} on the vehicle during a stand-alone ABS braking control maneuver on a split-μ road surface.

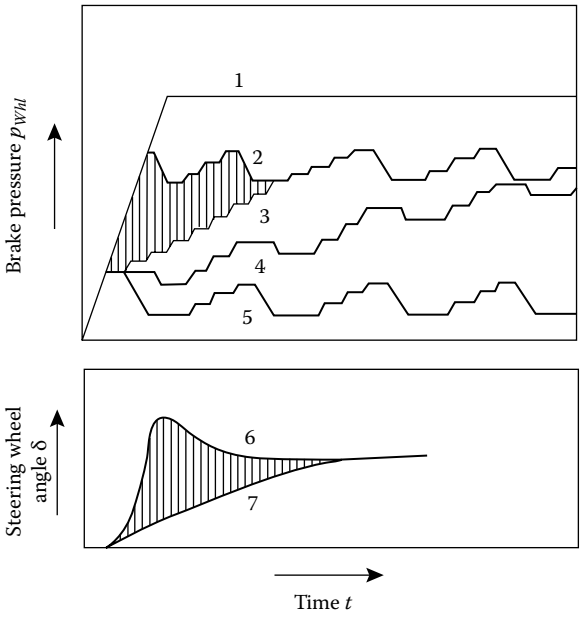


FIGURE 30.13 Influence of some yaw moment gradient reduction schemes on the required steering wheel angle effort. (1) Brake pressure induced by the driver, (2) individual wheel brake control, (3, 4) two possible reduced brake pressure gradient control schemes of the wheel on the high-μ surface, (5) brake control at the low-μ surface, (6) required steering wheel angle time history for individual control, (7) required steering wheel angle time history for reduced brake pressure gradient control.

have a negative effect on the stability of the car while braking in a turn. During cornering, the wheels on the outside of the turn have vertical forces which are larger than those on the inside of the turn, so that for individual control, the brake pressure at the wheels on the outside of the turn will be larger than those on the inside of the turn. "Braking while cornering" is therefore similar to "braking on split-μ." At the beginning of ABS control, the GMA feature will control the brake pressure to start with select low at the front and at the rear axle. This situation resembles partial braking in a turn from which it is known that the vehicle will tend to oversteer. The tendency disappears at the time that the wheels at the front axle are individually controlled (Figure 30.14). With a negative scrub radius, the car will also tend to oversteer during braking in a turn, since the front wheels will steer more into the turn. However, the situation is more severe, since the self-steering effect remains active during the entire maneuver. Therefore, the GMA feature must not be optimized for split-μ braking only, but a compromise must be found between vehicle controllability during braking on split-μ and while cornering.

On snow-covered roads, the μ-slip curve may not exhibit a maximum for slip values smaller than 100% (Figure 30.6). In this case and in general in all cases where the coefficient of friction continuously increases with the tire slip (e.g., deep snow, deep sand), the basic assumption of ABS (a maximum of the μ-slip curve exists for small slip values where the lateral stability margin is large) is violated. For these cases, ABS control may still eventually lead to locked wheels.

The hydraulic principle of pressure modulation for one wheel is shown in Figure 30.15. The components in the hydraulic block diagram are shown in the untriggered position, the inlet valve (EV) is then open, and the outlet valve (AV) is closed. If the driver applies the brake pedal, the brake fluid is pressed from the brake master cylinder (MC) through the open EV to the brake wheel cylinder (BC). If the EV is triggered, the brake pressure in the brake BC is kept constant. If the AV is triggered, then the brake fluid can escape from the brake BC to the low-pressure accumulator (Spk), and the pressure in the brake BC is reduced. If the motor (EM) is triggered, the recirculation pump (RFP) transports the brake fluid from the low-pressure accumulator back to the brake MC.

The block diagram of the hydraulic system for ABS is shown in Figure 30.16. Here, the EVs show an additional check valve. This check valve is required to release the brake fluid immediately from the brake BCs if the driver releases the brake pedal while the control is in operation. Furthermore,

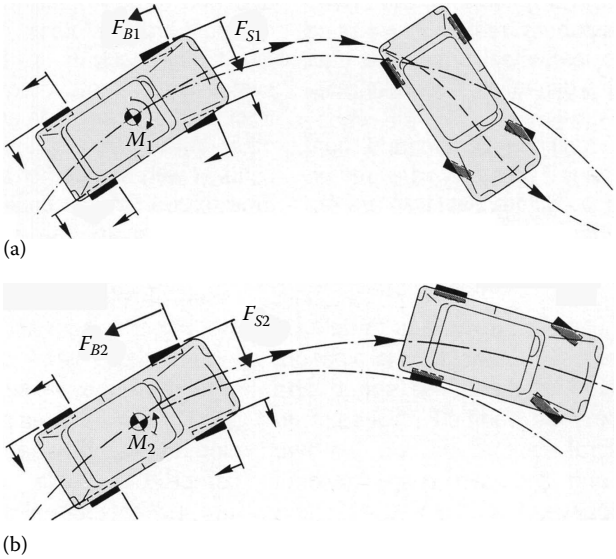


FIGURE 30.14 Influence of the yaw moment gradient reduction on the yaw moment on the car during ABS braking in a turn. (a) Oversteer behavior due to yaw moment gradient reduction and the resulting oversteering yaw moment on the car M_1. (b) Understeer behavior due to individual wheel brake control and the resulting understeering yaw moment on the car M_2.

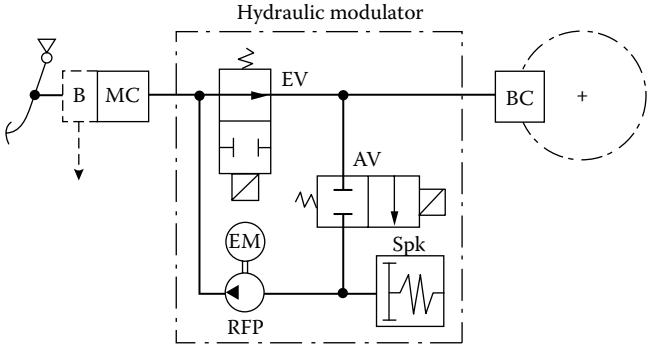


FIGURE 30.15 Principle of brake pressure modulation of one wheel brake cylinder.

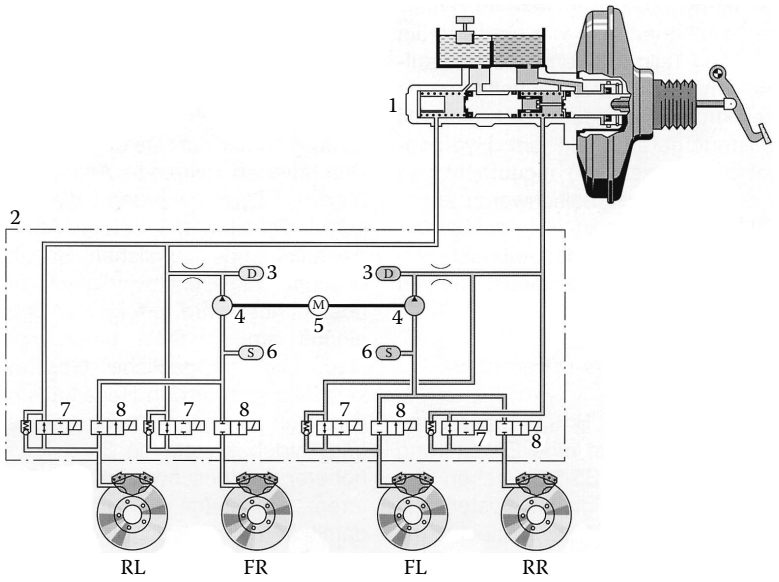


FIGURE 30.16 Block diagram of a typical hydraulic system for ABS: (1) brake MC with brake pedal, brake booster, and brake fluid reservoir, (2) HU, (3) damping chamber, (4) recirculation pump, (5) pump motor, (6) low-pressure accumulator, (7) inlet magnetic valve, (8) outlet magnetic valve.

a damping chamber (D) is shown. This is a cavity in the hydraulic block with a connection to the outlet side of the recirculation pump and a connection to the brake MC. The connection to the brake MC has a small orifice. The combination of the chamber and the orifice damps the pump pulsations in the fluid pressure and thus reduces the pump noise.

The inlet and the AVs are shown in Figure 30.17. The fluid filters are required to keep particles away from the valve seat, otherwise the valve may leak. The check valve function of the EV is realized by the special sealing of the valve in the housing.

The block diagram of the electronic control unit (ECU) is shown in Figure 30.18a. Inputs to the ECU are the four signals from the wheel speed sensors (WSSs). Outputs from the ECU are the stimulation voltages of the eight valves (9). The signals are processed (4) to derive the wheel velocities. The control logic is programmed in a microprocessor (6). For safety reasons, the ECU houses microprocessors which work in parallel. By continuous comparisons between the results from the microprocessors, the proper function of them is checked. If the results deviate for some time, a failure code is written to the EEPROM (8), the system is switched off (11), and the warning light is lit (13). With a diagnostics reader of the workshop (2), the failure can be read out. The ECU

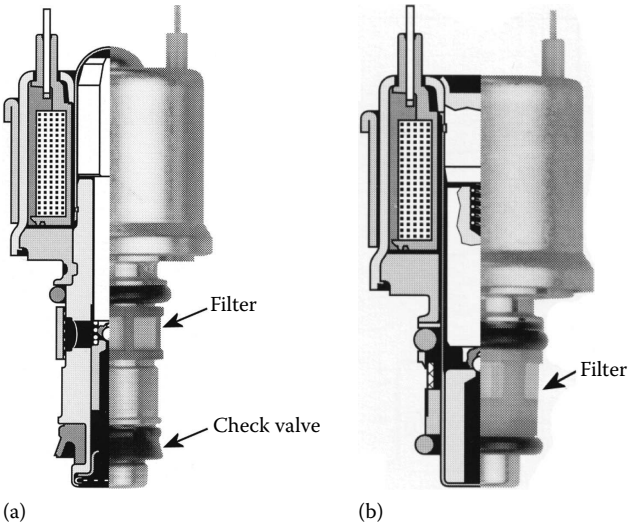


FIGURE 30.17 Valves used for ABS: (a) inlet magnetic valve and (b) outlet magnetic valve.

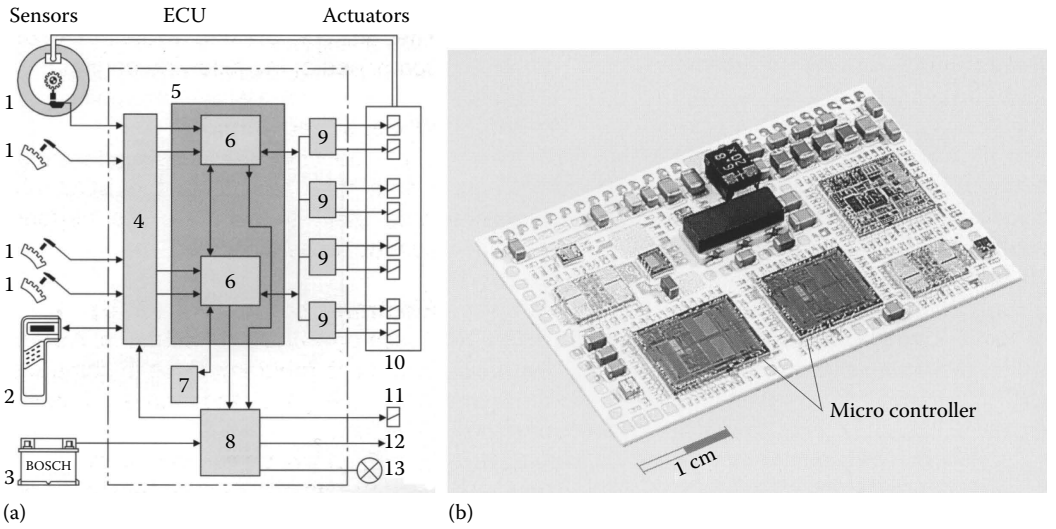


FIGURE 30.18 Block diagram of the ECU (a) and photo of a micro hybrid ECU (b). (1) WSS, (2) diagnosis reader, (3) battery, (4) signal processing, (5) digital controller, (6) microprocessor, (7) nonvolatile memory, (8) voltage stabilizer and failure memory, (9) output drivers, (10) magnetic valve coils, (11) valve and pump relays, (12) stabilized voltage supply, (13) ABS warning light.

is attached to the valve block of the hydraulic unit (HU). Since the HU is often mounted under the hood, near the engine, the ambient temperature can be high and specified to be up to 120°C. Figure 30.18b shows the ECU as a micro hybrid with the redundant microprocessors.

Figure 30.19 shows the complete HU with the attached ECU. The HU is mounted in a bracket with rubber mounts which is attached to the chassis of the car.

For the measurement of the wheel velocity, a WSS is used (Figure 30.20). The gear ring rotates with the wheel, while the sensor element is attached to the wheel carrier. The sensor works like an AC generator. It has a permanent magnet and a pole pin. If the gear ring rotates, then the magnetic flow from the magnet through the pole pin to the ring varies. The flow is high if a tooth of the ring is opposite to the pin and low between the teeth. Because of the varying magnetic flow, an alternating

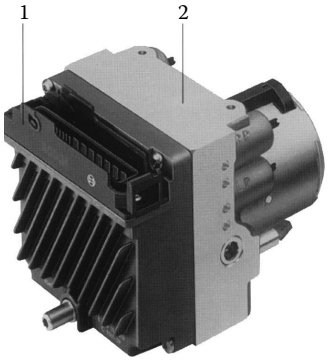


FIGURE 30.19 HU with attached ECU (1) and hydraulic block (2).

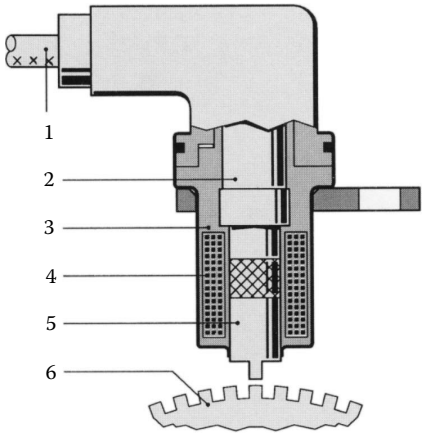


FIGURE 30.20 Typical passive WSS: (1) electrical cable, (2) permanent magnet, (3) housing, (5) electrical coil, and (6) gear wheel.

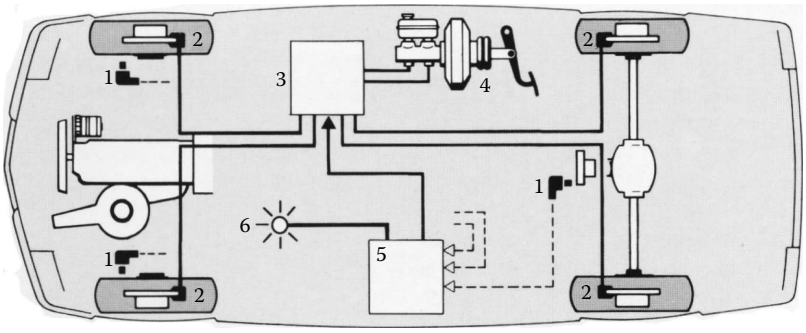


FIGURE 30.21 Integration of ABS in the rear wheel driven vehicle systems (1) wheel velocity sensors, (2) brake BCs, (3) hydraulic modulator, (4) brake MC with brake pedal and brake booster, (5) ECU, and (6) ABS warning light.

voltage is induced in the coil. The frequency of the alternating voltage is proportional to the wheel velocity. In the ECU, the periodical time of the alternating voltage is measured, transformed, and filtered to deliver the wheel velocity.

All components mounted in the car are shown in Figures 30.21 and 30.22. At the rear axle of the rear wheel driven car, a single WSS can be used (Figure 30.21), which picks up the average

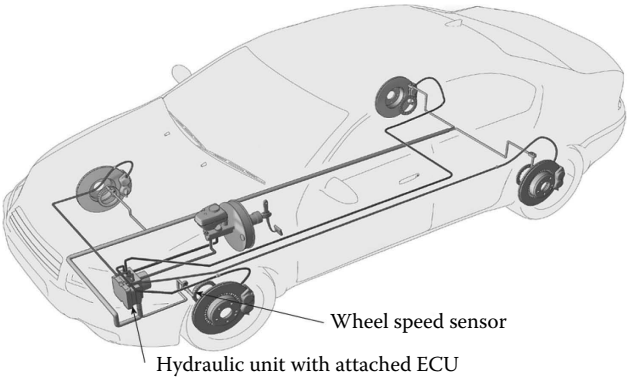


FIGURE 30.22 ABS mounted in the car.

velocity of the rear wheels at the drive axle (multiplied by the transmission ratio of the differential). This is possible in combination with select low control at the rear axle. By adaptation of the thresholds of the control logic, both wheels can be controlled on the basis of this single signal. If a system failure is detected, then the ECU must warn the driver by lighting the warning lamp (6).

30.1.4 Traction Slip Control System

Soon after ABS entered the automotive market, developments on the traction control system (TCS) started. The first implementations needed many expensive additional components like a high-pressure accumulator [7]. Since the main property of TCS for rear-wheel drive vehicles was seen as an improvement of vehicle stability, further developments of TCS concentrated on improved control of the engine output torque [8,9]. In order to avoid the cost of the "electronic accelerator pedal," simplified methods like "fuel injection suppression" were implemented to control the engine output torque [10]. These simplified methods have disappeared with the standard introduction of the electronic accelerator pedal on all vehicles.

If the driven wheels spin, the driver cannot control his car. The first requirement of TCS is to prevent the driven wheels from spinning. On the one side, the spin of the driven wheels can be controlled by controlling the engine torque. On the other side, wheel spin can also be controlled by active braking (i.e., the driver does not apply the brake) of the driven wheels. In general, TCS uses both controls. In principle, TCS controls the wheel slip to be close to λ_0.
The most important requirements of TCS are as follows:

1. Prevention of wheel spin and provision of track stability (neither drift nor swerve) while cornering.
2. Prevention of wheel spin on split-μ and slippery road surfaces and provision of sufficient traction.
3. Prevention of wheel spin at drive off uphill.
4. Priorities during handling:
   a. At high velocities:
      i. Vehicle stability
      ii. Comfort (noise, jerk)
      iii. Traction
   b. At low velocities:
      i. Vehicle stability
      ii. Traction
      iii. Comfort (noise, jerk)

5. Priorities during straight-line acceleration at low velocities:
   a. Traction
   b. Vehicle stability
   c. Comfort (noise, jerk)
6. Priorities during drive off and driving on split-μ:
   a. Traction
   b. Vehicle stability
   c. Comfort (noise, jerk)
7. Reduction of tire wear.
8. Reduction of the stress of the differential.
9. Warning to the driver if the control (control of the engine torque) is active.
10. Many of the ABS requirements are also valid for TCS.

The control concept of stand-alone ABS (control of the wheel acceleration and deceleration) cannot be used for the stand-alone TCS control. The reason is that by using the wheel acceleration and deceleration values, it is not possible to clearly differentiate between stable and unstable slip values. On the one side, the accelerations and decelerations are small in particular in the first gear of the transmission. On the other side, the engine torque is not independent of the engine rotational velocity. Therefore, it was necessary to find another control concept different from that used for ABS. Fortunately, it is possible to realize a slip control concept for TCS. The nondriven wheel velocities can be used to determine the free rolling wheel velocities and thus the drive slip of the driven wheels. However, this is not possible for four-wheel driven cars. With four-wheel driven cars, it is not possible for stand-alone TCS to keep the wheels from spinning.

TCS control can be basically split into slip control on homogeneous roads and on split-μ roads. On homogeneous roads, the coefficient of friction is the same at all driven wheels. Because of the differential, the drive torque is equally divided between the driven wheels at the drive axle. For straight-line maneuvers where the normal forces on both tires are equal, the drive slip is still the same at the two wheels. If the drive torque exceeds the maximum road torque, then both wheels on the axle will spin. In this case, it is sufficient to decrease the drive torque in order to reduce the slip at both wheels. This is realized by the engine management system. With gasoline engines, this is done by control of the throttle valve (Figure 30.23), and optionally, for fast torque control, by additional control of the spark timing

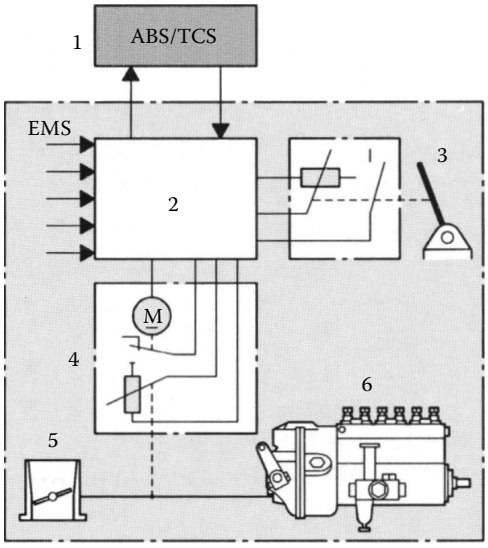


FIGURE 30.23 Electronic engine power control (EMS) with interface to ABS/TCS (1), ECU (2), accelerator pedal (3), stepped motor (4), throttle valve body of the gasoline engine (5), control rack of the injection pump of the diesel engine.

and fuel injection. With diesel engines, it suffices to control the control rack of the fuel injection pump. The controller (2) continuously sends the actual engine torque to the TCS. If the engine torque must be reduced, then TCS sends a reduced torque request to the controller. Since this request has a higher priority than the request from the driver (3), the torque is reduced to that requested by TCS. The actuator (4) of the controller then adjusts the throttle valve angle of the gasoline engine, respectively, and the rack of the diesel engine. Note that TCS can only reduce the engine torque from that requested by the driver (3).

On split-μ roads, the wheel on the low-μ surface (μ_l) tends to spin first if the drive torque at the cardan axle M_{Ca} is increased (Figure 30.24). A similar situation exists during cornering. The wheel on the inside of the turn tends to spin first and corresponds to the wheel on the low-μ surface, while the wheel on the outside of the turn corresponds to the wheel on the high-μ surface. Because of the differential, the drive torque on the wheel on the high-μ surface (μ_h) is limited to that of the wheel on the low-μ surface. The traction forces on both wheels (F_l) are correspondingly small. In order to achieve a larger traction force at the wheel on the high-μ surface, the wheel at the low-μ surface can be actively braked by the hydraulic system of TCS. The brake torque is controlled such that the slip velocity of the wheel on the low-μ surface lies below a velocity threshold which depends on the vehicle velocity. As long as the wheel on the high-μ surface does not spin, the engine torque is not reduced. However, if the slip velocity of the wheel on the high-μ surface also exceeds the velocity threshold, the engine torque will be reduced by TCS. Several slip velocity thresholds are used. The larger ones are used for the higher vehicle velocities. Their values are chosen such that a good compromise is reached between stability and traction during cornering and split-μ driving.

Sometimes the engine torque cannot be reduced fast enough to have a sufficient lateral stability margin at the driven wheel. In particular, for gasoline engines, pure throttle control can be too slow in decreasing the engine torque (Figure 30.25) and a large slip velocity results for a long time. In these cases, the control

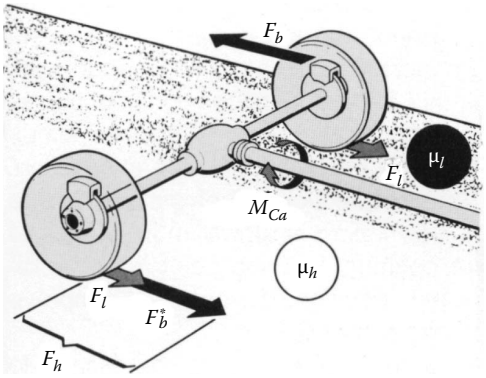


FIGURE 30.24 Principle of brake control for traction on split-μ road surfaces (μ_l is the low coefficient of friction and μ_h is the high coefficient of friction of the road surface): F_b is the force on the brake pad, F_b^* is the increment in the traction force, F_h is the enhanced traction force, and F_l is the maximum traction force on the slippery road surface.

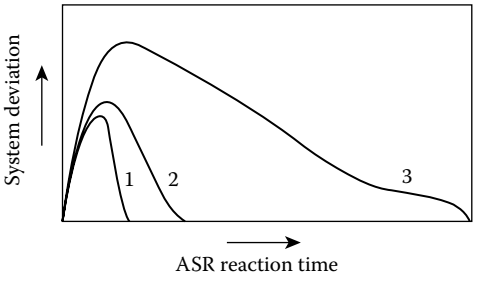


FIGURE 30.25 Influence of fast and slow interventions on the magnitude and the duration of the actuating signal of the TCS engine control: (1) fast brake intervention using a high-pressure accumulator, (2) fast engine intervention by spark-timing control, (3) slow engine intervention by throttle control.

of the spark timing and fuel injection can improve the situation. If a fast active brake system like the electrohydraulic brake system is available, the brakes can also be used to decrease the slip velocities at the driven wheels. In general, the brakes, the spark timing, and the fuel injection are only used to control the fast part of the drive torque reductions. They should only take care of the difference between the engine torque as a result of the throttle valve angle control and the required (nominal) engine torque. As soon as the required engine torque is reached, by throttle control, these fast parts of the control vanish. Since the control of the spark advance and the fuel injection has an influence on the exhaust gas emissions, there is a tendency to limit these control interventions to certain time slots or to forbid them all.

It is important to keep the difference in the velocities between the driven wheels small. The larger the velocity difference, the faster the differential rotates. Since the differential is also heavily loaded by the brake torque of the wheel on the low-μ surface, the power dissipated at the differential increases as much as the differential velocity increases. This dissipation may damage or even destroy the differential which is not designed for such large power dissipations.

On snow-covered roads, the wheel must not spin too long at the same spot since then the snow may get polished and turn into ice whose coefficient of friction is much lower than that of snow. Also, on deep sand tracks like in the desert, the wheels must not spin too long at the same spot since then they may dig holes into the sand and it is much more difficult to get out of the hole then to drive off on a smooth surface.

The use of the electronic accelerator pedal, EMS (Figure 30.23), allows also for the control of the slip of the driven wheels during lift-off of the accelerator pedal. If the accelerator pedal is released, then the engine torque acts like a brake torque at the driven wheels (engine drag torque). Particularly, on road surfaces with a low coefficient of friction, the slip of the driven wheels may become large because of the engine drag torque. In order to reduce this slip, the engine rotational velocity must be increased. This increase is realized in a special controller, the "Engine Drag Controller," whose acronym is MSR. In the MSR case, it is allowed to request an engine torque which is larger than that requested by the driver. If the MSR increases the engine rotational velocity, then the drive slip of the driven wheels is increased (the absolute value is decreased) and the side force reduction due to slip is reduced. Thus, the lateral stability margin of the driven wheels is increased.

The components of TCS are the same as those used for ABS. However, each of the driven wheels must have an individual WSS in order to allow for individual wheel brake control. The HU must be modified and extended in order to actively brake each of the driven wheels individually (Figure 30.26). The most important additions are two new valves in each brake circuit (3, 4). The most important modification is that the recirculation pump (7) is self-priming and can suck brake fluid through the brake MC from the brake fluid reservoir (1). During TCS operation, the suction valve (3) is opened and the pump supplies fluid to the brake BCs of the driven wheels. This valve has a large seat diameter so that the flow resistance for the brake fluid is low (Figure 30.27). In order to prevent the brake fluid from flowing back to the brake MC, the active pressure valves (4) are closed. The pressure-limiting function of the active pressure valves (4) is required to prevent pressure in the HU from increasing so much that the EVs are forcefully opened or that the HU gets damaged. If the pressure in the HU reaches this limiting value, the stem of the valve moves against the pressure-limiting spring to open the valve.

The TCS components in the car are shown in Figure 30.28. If the engine output torque is controlled by the system, then the TCS information light starts to blink. In case of a system failure, the TCS light is permanently on.

30.1.5 Electronic Stability Program

The first investigation of a feedback control of the stability and steerability of a vehicle realized by control of the individual wheel slips was given in 1973, much before ABS entered the automotive market [11]. Industrial realizations were proposed in the late eighties [12–14]. These proposals also communicated the need of system integration in the future and considered active suspension and active steering control. The first realization of active rear wheel steering control entered the automotive market in 1991, where a yaw velocity sensor was used for feedback control of the yaw rate of

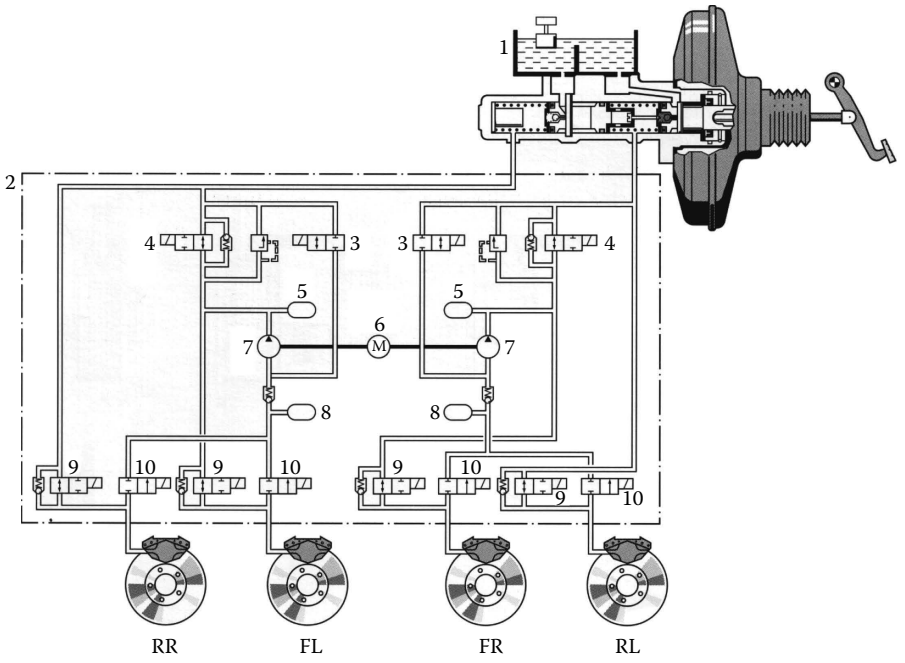


FIGURE 30.26 Block diagram of a typical hydraulic system for TCS: (1) brake MC with brake pedal, brake booster, and brake fluid reservoir, (2) HU, (3) suction magnetic valve, (4) active pressure magnetic valve with pressure-limiting function, (5) damping chamber, (6) pump motor, (7) self-priming recirculation pump, (8) low-pressure accumulator, (9) inlet magnetic valve, (10) outlet magnetic valve.

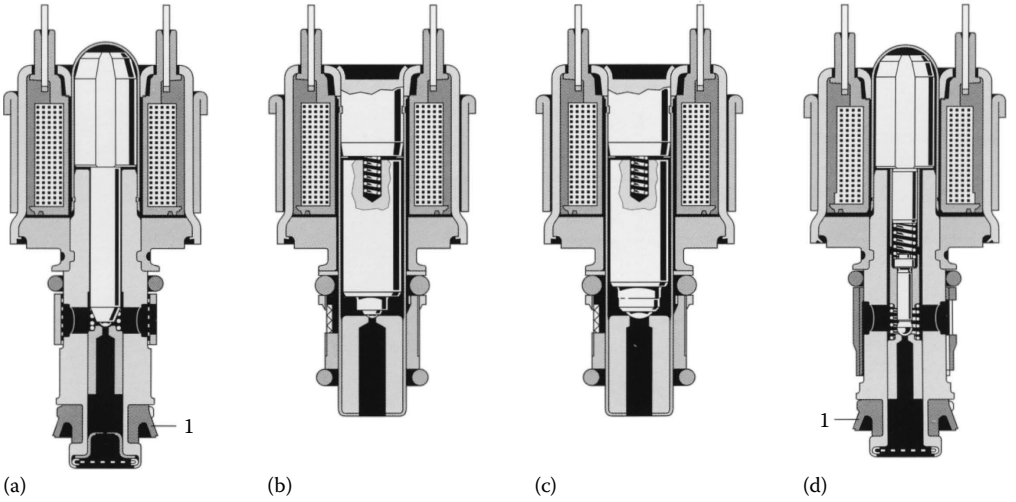


FIGURE 30.27 Valves used for TCS: (a) inlet magnetic valve, (b) outlet magnetic valve, (c) suction magnetic valve, and (d) active pressure magnetic valve, (1) check valve function integrated in lip seal.

the car [15]. Other realizations followed where the feedback control of the yaw rate was realized by control of the brake or traction force distribution [16–18].

ESP entered the automotive market in 1995 [19,20]. The system uses a hierarchical control structure and will be discussed in this paragraph. With rapid succession, further systems entered the market [19–31]. Many of these systems used the stand-alone ABS and TCS controller structures to

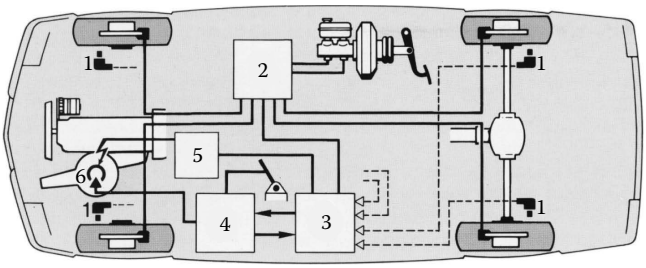


FIGURE 30.28 Integration of TCS in the vehicle system: (1) wheel velocity sensors, (2) hydraulic modulator, (3) TCS ECU, (4) electronic engine power control, (5) electronic injection and spark control, and (6) throttle actuator.

realize ESP and are called modular systems (see e.g., [29]). Further developments' main goal was to improve system reliability and reduce system cost [32,33]. A new proposal by Toyota, called VDM (Vehicle Dynamics Management), to fill the gaps between the different controller parts of the modular approach is based on a hierarchical controller structure and includes tire force signals [34].

Car swerve and drift are described by the body slip angle. The slip angle of the car has also a large influence on the controllability of the car by the driver and on the vehicle stability as described earlier. Therefore, in order to keep the car stable, the slip angle must be limited and also kept small so that the car behaves as usual and the driver can control his car. In this respect, it is useful to formulate the usual behavior of an average driver on dry roads during normal driving.

- The average driver steers such that the lateral acceleration is usually less than 0.2 g.
- The average driver drives such that the slip angle of the car is usually less than 2°.
- The average driver applies the brake pedal such that the vehicle deceleration is usually less than 0.4 g (brake pressure about 40 bar).

The average driver therefore is used to drive in the linear region of the slip and the slip angle curves of the tires where the car behavior is linear. His experience in the nonlinear region of the car behavior is virtually zero. He has no idea of the coefficient of friction of the road and of the stability margin of the car [35]. If the car behaves nonlinearly and its stability margin is small, the driver is caught by surprise by its substantially different behavior. Often he panics and does not act thoughtfully but, for example, steers too much. It is therefore not legitimate to expect thoughtful behavior of the driver in the nonlinear region of the car behavior while designing active safety systems like ESP. In order to avoid panic by the driver, it seems to be a good idea to start out with a nominal behavior of the car which is close to linear.

ESP is designed to improve the track stability of the car in every driving situation. However, it is not designed to control the brakes and the engine for the improvement of the basic handling behavior of the car. Therefore, ESP must start the control only if the physical limit of the car motion is reached. As ESP was installed on cars which may tow large trailers, the control has been extended to stabilize such combinations if the trailer starts to sway. Furthermore, as ESP also gets installed on vehicles with a soft suspension and with a high center of mass like on off-road vehicles and on Sport Utility Vehicles (SUV), its capabilities have been expanded to also improve the rollover stability.

30.1.5.1 System Concept

The first requirement on ESP is to keep the slip angle of the car limited to small values. From the discussion on vehicle controllability, it is important to keep the slip angle small and to keep the combination $(\beta, \dot{\beta})$ in the stable area. A sensor for the slip angle would allow for an easy formulation of the actuating signal. However, such a sensor for automotive applications (reliability at all

weather conditions, cost) is not available. As an alternative, a yaw rate sensor in combination with a lateral acceleration sensor is used. The derivation of the slip angle from these sensor signals is not always sufficiently reliable and a second actuating signal is required based on the nominal yaw velocity of the car. For the derivation of the nominal yaw velocity, the steering angle at the front wheels, the vehicle velocity, MC brake pressure, and the engine torque is required. Thus, the resulting system concept requires, besides the WSSs, a yaw rate sensor also, a lateral accelerometer, a steering wheel sensor, and a brake pressure sensor. The engine torque value is supplied by the engine management system. Figure 30.29 shows the system concept.

Since the vehicle moves at the physical limit, it is not possible to increase the side forces on the wheels and thus reduce the slip angles without influencing the longitudinal forces. Instead, ESP controls the yaw motion to indirectly control the slip angles. Thus, ESP computes the required yaw moment from the actuating signals. Increments in the yaw moment can be generated by modifying the lateral and the longitudinal forces on the tires. However, with the brakes alone, the slip angles of the tires cannot be directly modified. Indirectly, the lateral forces can be modified by the tire slip as can be seen from Figure 30.9, but not independent from the longitudinal force. Both forces have to be considered if a change in the slip is made to generate the required yaw moment. As an example, the yaw moment due to a change in the longitudinal slip of the left front wheel ($\Delta\lambda_1$) is (Figure 30.31).

$$\Delta M_{Yw} = -\frac{\partial F_{S1}}{\partial \lambda_1} \cdot \Delta\lambda_1 \cdot (a \cdot \cos\delta_W - b \cdot \sin\delta_W) + \frac{\partial F_{B1}}{\partial \lambda_1} \cdot \Delta\lambda_1 \cdot (b \cdot \cos\delta_W + a \cdot \sin\delta_W) \qquad (30.2)$$

If the partial derivatives in Equation 30.2 are known, then the slip increment for each wheel can be computed using, for example, the steepest descent method. The partial derivatives represent the slopes of the brake and side force coefficient curves in Figure 30.9 for a constant wheel load. Since they are unknown, vary rapidly with the slip value, and may vary with the tire make and state, they may be computed by fast identification. However, for fast identification, the (pressure) input should be rich enough, for example, a pressure pulse or a pressure step. But then transients in the wheel dynamics are excited and the result of a stepwise increase in the brake pressure is shown in Figure 30.30 [36]. This shows that the slope of the steady-state brake force–slip curve of Figure 30.9 cannot be determined by

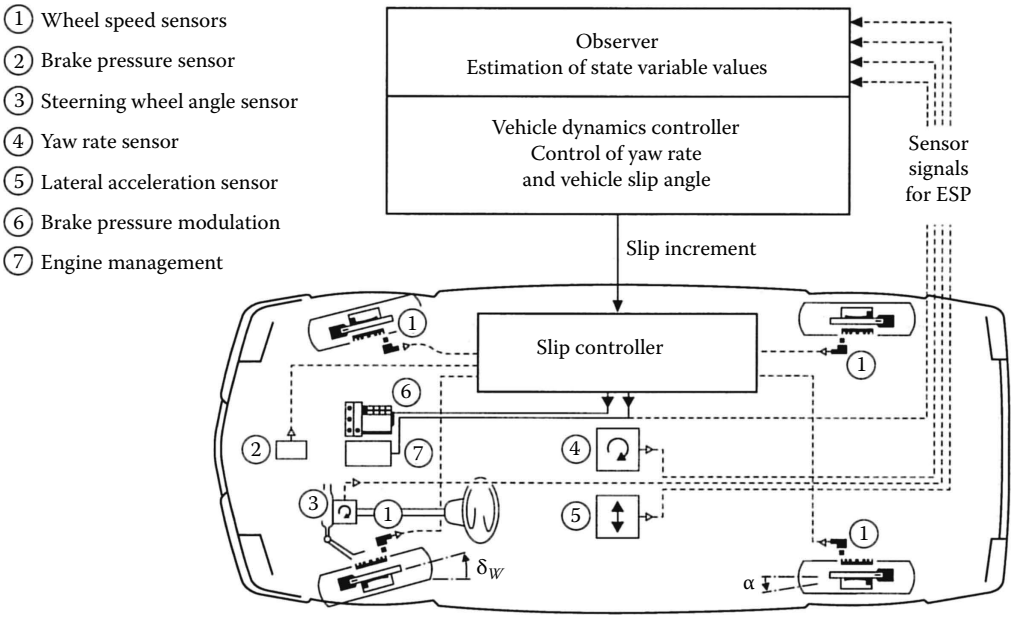


FIGURE 30.29 System concept of ESP with hierarchical control structure.

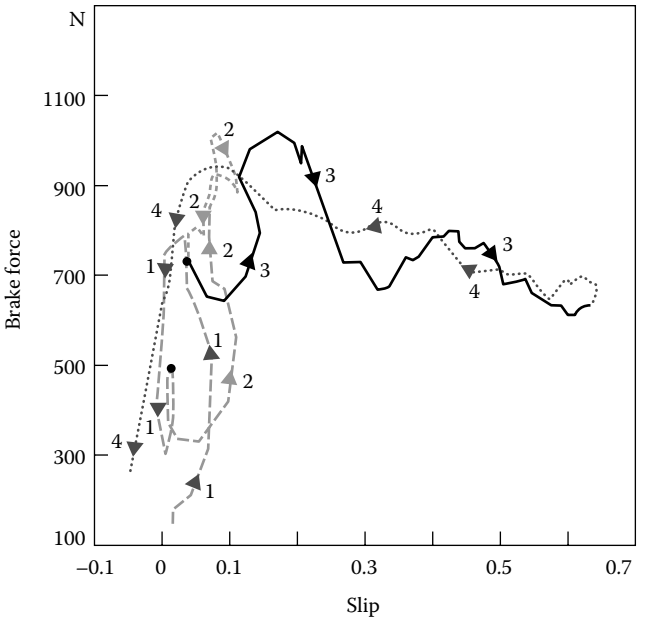


FIGURE 30.30 Dynamic brake force—slip phase diagram where the brake torque is increased in a stepwise manner (from number 1 to number 3) and decreased (number 4).

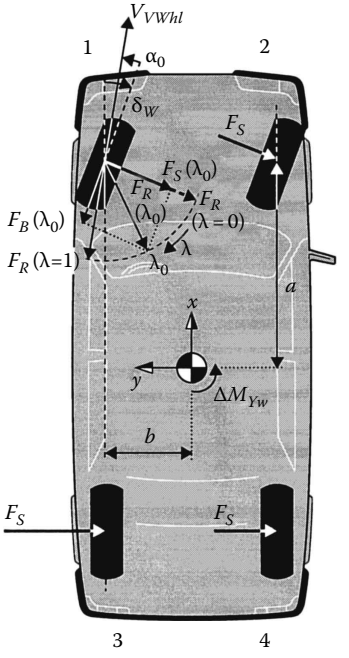


FIGURE 30.31 Influence of brake slip modification on the yaw moment on the car.

identification without consideration of wheel dynamics. The identification then becomes complex with many unknown variables and is no longer practical.

Instead, the friction ellipse concept is used (Figure 30.31). The assumption is that the friction force between the tire and the road has reached its maximum value (i.e., the car motion has reached its physical limit). For simplicity of explanation, the friction ellipse is approximated by a friction

circle. If the wheel slip is zero, then there is no longitudinal force and the yaw moment of the side force of the left front tire, wheel number 1, is with $F_{S1} = F_{R1}(\lambda = 0)$

$$M_{Yw} = -F_{S1} \cdot (a \cdot \cos\delta_w - b \cdot \sin\delta_w) \tag{30.3}$$

If the wheel slip is increased to λ_0, then a corresponding brake force ($F_{B1} = F_{B1}(\lambda_0)$) results and the side force is reduced to $F_{S1}(\lambda_0)$. The vectorial sum of the side and the brake force is denoted as $F_{R1}(\lambda_0)$, which has the same magnitude as the initial side force because of the preceding assumptions. Since F_{R1} has a direction which is different from that of the side force F_{S1}, the yaw moment on the car has changed due to the increment in the wheel slip (in the explanation the increment is the change in slip from 0 to λ_0). The gain (yaw moment increment due to an increment in the tire slip) is determined using the friction ellipse for which knowledge of the slip is required. One limiting condition is that the brake slip can only be increased up to the point where the wheel locks. Another limiting condition is that the engine torque cannot be increased beyond that value induced by the driver (except if the drag torque of the engine is too large such that large slip values result). During partial braking, however, increase in the brake pressure is allowed at any wheel beyond that value induced by the driver.

The change in the yaw moment is therefore controlled by changes in the wheel slips, which can be calculated from the friction ellipse using the wheel slips. The slip controller, which is below the vehicle dynamics controller in the hierarchical control system, can do both tasks. It computes the wheel slips and controls the wheel slip increments. Since by design, the stand-alone ABS is not able to control brake slip increments and to determine the wheel slip, the brake slip controller (BSC) of ESP is designed differently (for a different ESP design based on stand-alone ABS and TCS see [29]). A special feature of the slip controller, which computes the target slip λ_T which is close to the maximum of the μ-slip curve, allows it to be used for the ABS-function. With some modification, this special feature is also used for the TCS function of the slip controller. Furthermore, the slip controller is also used for all other driving situations such as coasting and partial braking to control the slip increments as commanded by the superimposed vehicle dynamics controller. It is important to notice that the structure of the underlying slip controller with its ABS and TCS functions is independent of the interventions of the superimposed vehicle dynamics controller.

In the stable area of the μ-slip curve, a unique relation exists between the wheel slip and the brake and drive torques. Here, nominal slip values can be replaced by nominal torque values and the brake and engine torque may be controlled to this nominal value. Care must be taken, however, that during this control to a nominal torque, the wheel does not lock nor spin, for example, because of sudden changes in the longitudinal tire force value. For slip values in the unstable area, this control to a nominal pressure is not possible. For improved yaw control performance, the control to a nominal wheel torque is used for very small nominal slip values. The reason here is that the μ-slip curve is very steep and small estimation errors in the slip result in large errors in the brake and drive torques from their nominal values. The decision to control the slip to the nominal slip value or to control the wheel torque to the nominal torque value is taken in the slip controller.

The hierarchical algorithm will be discussed using Figure 30.32 (see also [34] for a hierarchical design of ESP). On the top left side of the figure, the driver's desire in terms of the desired vehicle motion is derived from his inputs, while on the top right side, the actual motion of the car is measured and motion variables that are not measured are estimated.

The desired vehicle motion is derived using the linear steady-state bicycle model. From this model, the first nominal values of the yaw velocity $\dot{\psi}_M$ and the slip angle β_M are obtained, using the vehicle velocity. Since the lateral acceleration of the bicycle model is limited by the coefficient of friction of the road, the yaw velocity has to be limited also. At high vehicle velocities, the allowed slip angle is reduced in order to give the driver better support in controlling his car. These desired motion values must be compared with the actual motion values of the car to deliver the actuating signals. The actual yaw velocity is measured so that only the actual slip angle must be estimated. This is done by an observer

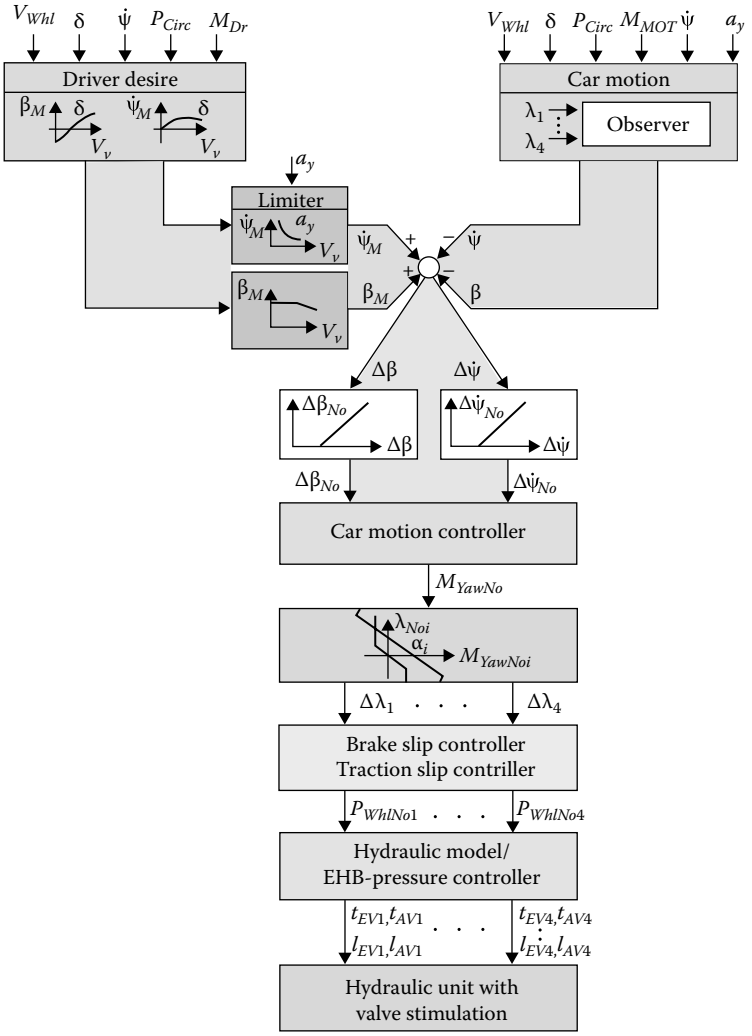


FIGURE 30.32 Control blocks of ESP.

who, among others, needs the actual slip values of the tires. The deviations between the desired and the actual motion values are filtered by a dead zone in order to avoid interventions which result, for example, from modeling, measurement, and estimation errors. The resulting filtered deviations are the actuating signals for the car motion controller. Using a simple proportional-integral-differential-type controller (PID) controller from the actuating signals, the required yaw moment on the car is determined. The gains are variable and depend among others on the magnitude of the slip angles of the tires.

In the next lower block, the nominal value of the slip for each wheel is computed using the friction ellipse (30.12) in order to achieve the desired yaw moment on the car. Usually, the slip is modified only at two wheels. For instance, during full braking in a turn if the vehicle shows an incipient swerve and the motion has to be stabilized, the slip at the front wheel on the outside of the turn is increased while the slip at the rear wheel on the inside of the turn is decreased. On the other side, if during full braking in a turn the drift has to be reduced, the slip at the front wheel on the outside of the turn is reduced while the slip at the rear wheel on the inside of the turn is increased. If the brake slip of the rear wheel on the inside of the turn is zero, it is not possible to reduce it further. Therefore, if the coasting vehicle has to be stabilized in a turn, the brake slip on both wheels on the outside of the turn is increased. If the vehicle has to be stabilized during partial braking in a turn,

the slip at both wheels on the outside of the turn may be increased while the slip at the front wheel on the inside of the turn may be decreased. Therefore, the choice of the wheels used for the control of the yaw moment on the car depends on the driving situation. The difference between the desired slip value λ_{Noi} of wheel number i and the target slip value λ_{Ti} that is determined by the slip controller is sent to the slip controller as a desired increment in the tire slip $\Delta\lambda_i$.
The slip controller consists of two parts:

- The BSC. In this controller, the positive slip values are controlled as they occur during braking and during engine drag. Also, the ABS function is implemented and improved as compared with the stand-alone ABS by utilizing all additional sensor signals which are used for ESP. The target slip λ_{Ti} for ABS is tuned for each wheel number i such that a good compromise exists between stopping distance and vehicle controllability.
- The traction slip controller (TSC). In this controller, the negative slip values are controlled. Also, the TCS function is implemented and in particular for four-wheel drive vehicles improved as compared with the stand-alone TCS, by utilizing all additional sensor signals which are used for ESP. The target slip λ_{Ti} for TSC is tuned for each driven wheel number i such that a good compromise exists between traction and vehicle controllability.

The slip controller determines the target slips λ_{Ti} for the ABS and the TCS functions according to the coefficient of friction of the road. In general, the target slip increases as the coefficient of friction of the road increases. Simple PI and PID controllers are used in the slip controller to compute the required brake and traction torques on the wheels. While the traction torque can directly be requested from the engine management, the required brake torque must be transformed to valve stimulations of the hydraulic block. If the system includes an electrohydraulic brake system (EHB) [37], then the pressure request can directly be sent to the electronic pressure controller. Otherwise, the valve stimulations are computed using a hydraulic model of the actuator.

30.1.5.2 Performance Requirements

The requirements of ESP relate to the behavior of the car at the physical limit where the resultant tire forces are near their maximum values. At the physical limit, the longitudinal and lateral forces at the wheels depend on each other as shown by the friction ellipse. Therefore, the car behavior results from a compromise between the longitudinal and the lateral forces on the wheels. The car behavior is judged by expert drivers. Their judgment is based on measurements but their subjective judgment plays a dominant role. The subjective judgment relates to the compromise. Therefore, the requirements refer mainly to this compromise.

30.1.5.2.1 General Requirements

- The driver must feel safer with the system than without the system (i.e., system failures must not lead to dangerous vehicle behavior).
- The driver must feel safe with respect to the vehicle behavior (plausibility of the interventions).
- ESP must improve the track stability with respect to understeering and oversteering in all driving situations like braking and accelerating, at all road friction levels, and in particular on split-μ surfaces and at μ-jumps.
- ESP must perform on all road types, uphill and downhill, and on banked roads.
- ESP must keep the vehicle stable and steerable in all driving situations if the steering angle is continuously increased to large values or continuously decreased from large values.
- ESP must reduce the steering effort of the driver.
- The driver must not have the impression that with ESP the vehicle is slower than without ESP.
- Inputs from the driver must not lead to instability of ESP.

- ESP must be able to keep the slip angle of the vehicle smaller than a certain value, which may be reduced depending on the vehicle velocity, lateral acceleration, friction coefficient of the road, and steering wheel motion. If that is not possible, ESP must limit the slip angle velocity to such values which can be managed by the driver to keep the vehicle stable.
- The vehicle must react immediately to the steering input of the driver in order to keep the driver from steering too much. The settling (oversteering) intervention to reduce drift must not be too violent since the future driving situation cannot be predicted.
- ESP must immediately recognize the recovery to a stable vehicle motion.
- ESP must be able to deal with the kinematics and with the tolerances of the steering column.
- ESP must not intervene in banked turns on public roads (bank angle <20°) if the physical limit is not reached.
- ESP must not degrade the vehicle behavior in any circumstance (e.g., flat tire, trailer, bad shock absorbers).
- It is not required that ESP should work if the vehicle is not driving forward. Reverse driving must, however, be definitely recognized.
- ESP must show full performance on elevated roads up to 2500 m above sea level.
- The driver must maintain the possibility to drive his vehicle close to the physical limit without ESP interventions. Otherwise, he may complain about too early ESP interventions.
- The TSC must be tuned for an optimal compromise between traction and vehicle controllability such that the ESP interventions, which are annoying because they decelerate the vehicle, are minimized.
- The requirements of ABS and TCS also apply to ESP.
- The application effort of ESP should be as small as possible. The application should be possible using a minimum number of parameters. The controller structure should be clear and simple, otherwise the application effort increases. The application effort increases if various parameters cannot be independently applied and controller functions are not independent of each other.

30.1.5.2.2 Requirements during Full Braking (ABS Function)

1. Priorities during handling, high velocities, and full braking:
   a. Vehicle stability: $\beta < 5°$
   b. Braking distance: the braking distance must not increase because of the vehicle dynamics control
   c. Comfort (noise, brake pedal pulsation)
2. Open loop jump in steering angle during full braking:
   Within the first 3 s, the slip angle value must not exceed 6°.
3. During split-μ braking with friction coefficients 1.0 on the one vehicle side and 0.1 on the other side, the deceleration must not be smaller than 0.3 g.

30.1.5.2.3 Requirements during Partial Braking

- The vehicle deceleration must follow the pressure in the brake MC.
- The electronic brake force distribution (EBD) control must be tuned such that a minimum of ESP interventions is required.

30.1.5.2.4 Requirements during Traction Control (TCS Function)

Priorities in handling at high velocities:

1. Vehicle stability: $\beta < 6°$
2. Comfort (noise, vehicle jerk)
3. Traction

30.1.5.2.5 Miscellaneous Requirements

1. Closed-loop lane change while coasting:
   a. Slip angle $\beta < 5°$
   b. Angular speed of the steering wheel:
      i. On low μ: < 300°/s
      ii. On high μ: < 400°/s
2. Open-loop cornering: The lateral displacement of the vehicle must be less than 0.5 m (relative to the middle of the track).
3. In case of understeering, the vehicle may be decelerated.
4. The "Moose Test" (also called the "VDA-Test") must be passed [1].

30.1.5.3 Vehicle Dynamics Controller

The two main parts of the vehicle dynamics controller are the determination of the nominal yaw velocity and the determination of the vehicle slip angle. These parts will be discussed in the following paragraphs.

30.1.5.3.1 Computation of the Nominal Yaw Velocity

The nominal value of the yaw velocity is derived primarily from the linear steady-state bicycle model (Figure 30.33). The equations for this model are shown in Table 30.1.

The total lateral stiffnesses at the front and at the rear wheels, $C_{\alpha,F}$ and $C_{\alpha,R}$, respectively, include influences from the tires and from the compliance of the suspension and the steering system. The quotient $(a + c)/R_T$, in which R_T is the radius of the turn, is called the Ackermann steering angle δ_A. If the slip angle at the front axle is equal to that at the rear axle, then the steering angle is equal to the Ackermann steering angle. The vehicle handling behavior is then called "neutral steer." If the slip angle at the front axle is larger than that at the rear axle, then the vehicle handling behavior is called "understeer." If the slip angle at the front axle is smaller than that at the rear axle, then the vehicle handling behavior is called "oversteer." It can be easily shown that, for small vehicle slip angles, the product of the vehicle velocity and the yaw velocity equals the lateral acceleration. By definition, the characteristic velocity v_{CH} depends on the lateral tire stiffnesses, and also on the vehicle mass,

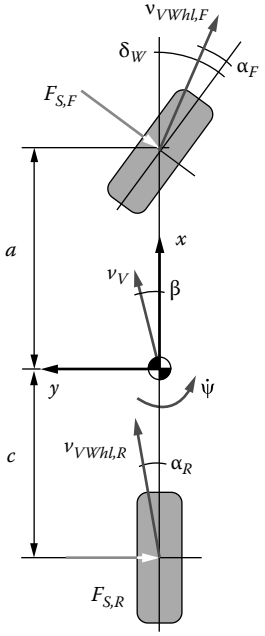


FIGURE 30.33 The bicycle model of the vehicle.

TABLE 30.1
Basic Equations and Results of the Linear Bicycle Model

Basic Equations	Results
Kinematic relationship	Yaw velocity
$\delta_W = \frac{a+c}{R_T} + \alpha_F - \alpha_R = \delta_A + \alpha_F - \alpha_R$	$\dot{\psi} = \frac{v_V}{(a+c)\cdot\left(1+{v_V}^2/{v_{CH}}^2\right)}\cdot\delta_W$
$\beta = \frac{c}{R_T} - \alpha_R$	Yaw gain
	$\frac{\dot{\psi}}{\delta_W} = \frac{v_V}{(a+c)\cdot\left(1+{v_V}^2/{v_{CH}}^2\right)}$
Linear tire relationship	
$F_{S,F} = C_{\alpha,F}\cdot\alpha_F, \quad F_{S,R} = C_{\alpha,R}\cdot\alpha_R$	$\frac{1}{{v_{CH}}^2} = \frac{m_V}{(a+c)^2}\cdot\left(\frac{c\cdot C_{\alpha,R} - a\cdot C_{\alpha,F}}{C_{\alpha,F}\cdot C_{\alpha,R}}\right)$
Dynamic equation	
$F_{S,F} + F_{S,R} = m_V\cdot v_V\cdot\dot{\psi}$	Vehicle slip angle
Steady-state equations	$\beta = \delta_A\cdot\left(\frac{c}{a+c} - \frac{a}{(a+c)^2}\cdot\frac{m_V}{C_{\alpha,R}}\cdot{v_V}^2\right)$
$F_{S,F}\cdot a - F_{S,R}\cdot c = 0$	
$v_V = R_T\cdot\dot{\psi}$	

the wheel base ($a + c$), and on the location of the center of mass. If the vehicle behavior is understeer, then the characteristic velocity exists. For sports cars, the value is larger (e.g., 30 m/s) than for other cars (e.g., 20 m/s). If the vehicle behavior is neutral steer, the characteristic velocity is infinity, and if the vehicle behavior is oversteer, then the characteristic velocity is imaginary. By design, all cars are understeer in their handling behavior.

The relations between the yaw velocity and the vehicle velocity for some steering wheel angles are shown in Figure 30.34. The curves show an almost linear increase in the yaw velocity with increasing vehicle velocity for low vehicle velocities, but for larger velocities the increase is smaller and for very large vehicle velocities the yaw velocity decreases with increasing vehicle velocities. It can be shown that the vehicle velocity at the maximum of the curves equals the characteristic velocity and that at the maximum the yaw velocity is half the yaw velocity of the neutral steer car. Since the lateral acceleration equals the product of the vehicle velocity and the yaw velocity,

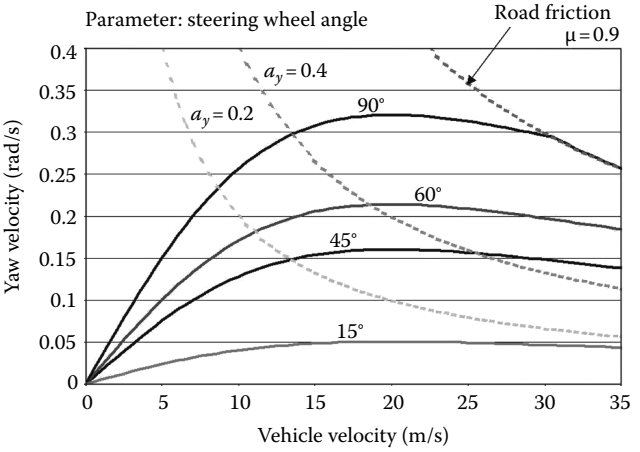


FIGURE 30.34 Handling characteristic of the linear steady-state bicycle model for some steering wheel angles.

curves of equal lateral acceleration are hyperbolas. If the vehicle velocity increases from zero, then the yaw velocity and the lateral acceleration also increase from zero. As shown for the steering wheel angle of 90°, the curve can cross the hyperbola of lateral acceleration of 0.9 g at a certain vehicle velocity. At this point, the lateral acceleration equals the coefficient of friction of the road. For larger vehicle velocities, the lateral acceleration according to the yaw velocity curve would be larger than the coefficient of friction of the road. This is not possible; hence, for larger vehicle velocities the yaw velocity is limited by the hyperbola representing the coefficient of friction of the road. The yaw velocity at a certain vehicle velocity increases proportionally with increasing steering wheel angle. Therefore, the yaw gain, which is the yaw velocity divided by the steering angle, is independent of the steering angle, that is, it is one curve for all steering angles.

The linear relationship between the side force and the slip angle holds only for small slip angles, that is, for small lateral accelerations. For larger slip angles, the relationship is nonlinear, so that different curves may be expected. Figure 30.35 shows the yaw velocity and the yaw gain curves for some steering wheel angles as measured in a front-wheel driven car. The curves show that the lateral acceleration of 1.0 g which corresponds to the coefficient of friction of the road is not reached for

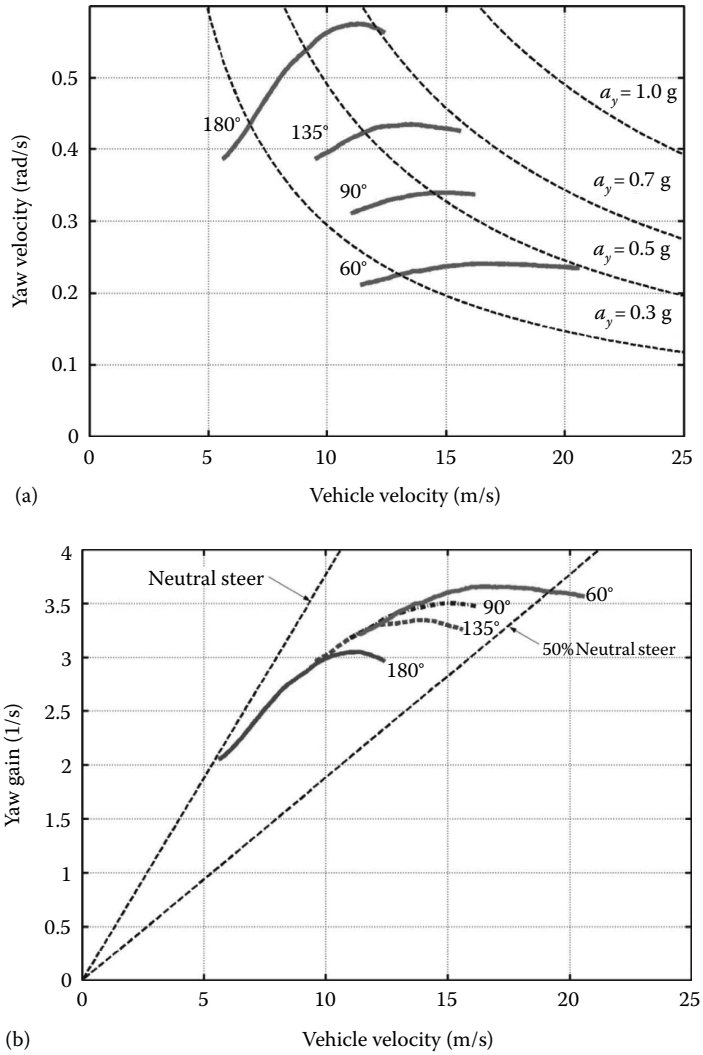


FIGURE 30.35 Handling characteristic of a front-wheel driven car for some steering wheel angles. (a) Yaw velocity, (b) Yaw gain.

any steering wheel angle because at the front axle the maximum of the μ-slip angle curve is reached while at the rear axle the maximum is not yet reached. The larger the steering angle, the smaller the vehicle velocity at the maximum of the yaw velocity curve. Contrary to the linear bicycle model, the yaw gain depends on the steering wheel angle and the maximum yaw gain decreases as the steering wheel angle increases. Therefore, the handling behavior of the linear bicycle model differs from that of the vehicle and the model must be extended to better represent the true handling behavior. Otherwise, the difference between the nominal yaw velocity computed by the model and the measured yaw velocity may become so large that an ESP intervention occurs although the physical limit is not reached.

The characteristic velocity v_{CH} varies with the vehicle mass (e.g., laden/unladen) and with the tire stiffness (e.g., summer/winter tires) and so also does the true handling behavior of the vehicle. Since these variations are unknown, ESP must be tuned to be robust enough to prevent interventions before the physical limit is reached. This robustness is reached by specifying a dead zone about the nominal yaw velocity as computed from the extended linear bicycle model.

As explained earlier, the lateral acceleration is limited by the coefficient of friction of the road surface (30.4).

$$\mu \geq |a_y| \tag{30.4}$$

Using (30.5) for a steady cornering motion, it follows that the yaw velocity must be limited by (30.6).

$$a_y = v_V \cdot \dot{\psi} \tag{30.5}$$

$$|\dot{\psi}| \leq \frac{\mu}{v_V} \tag{30.6}$$

As long as the coefficient of friction of the road is unknown, the lateral acceleration a_y is taken instead of μ. This choice will prevent ESP interventions before the physical limit is reached.

Before the yaw velocity computed from the extended bicycle model can be compared with the measured yaw velocity, it must be filtered to simulate the phase shift between the steering angle and the yaw velocity of the vehicle. Similarly, before the yaw velocity can be limited corresponding to the coefficient of friction of the road and the vehicle velocity, it must be filtered to simulate the phase shift between the yaw velocity and the lateral acceleration.

Two special situations need some attention. First, if the vehicle performs an ABS-braking maneuver on a split-μ road surface, then the driver has to countersteer to keep the vehicle on track. This countersteering must not be interpreted as a cornering desire from the driver. From the difference between the brake pressures at each axle, the yaw moment on the car is computed and the steering angle evaluated, which is required to countersteer against this yaw moment. This steering angle is used as an offset and the calibrated steering angle is used in the bicycle model to derive the nominal yaw velocity. Second, if the vehicle moves though a banked turn, then the gravitational acceleration has an influence on the measured lateral acceleration and the measured lateral acceleration may even become zero. Without special care, this situation may be interpreted as a vehicle which is swerving on an icy road surface. Since from the available sensors the situation at hand cannot be detected (swerving on ice or stable driving in a banked turn), a plausibility check is made. If the driver countersteers, then the situation is interpreted as swerving on ice and the driver is supported by ESP to stabilize the vehicle. If the driver does not countersteer, then the situation is interpreted as stable driving in a banked turn. In the latter situation, offsets for the lateral acceleration and for the steering angle may be computed and the sensor signals may be calibrated with these offsets. Further ESP control is then on the basis of these calibrated signals. As an alternative to the compensation of the sensor signals, it is also possible to increase the dead zones for the actuating signals of the vehicle dynamics controller.

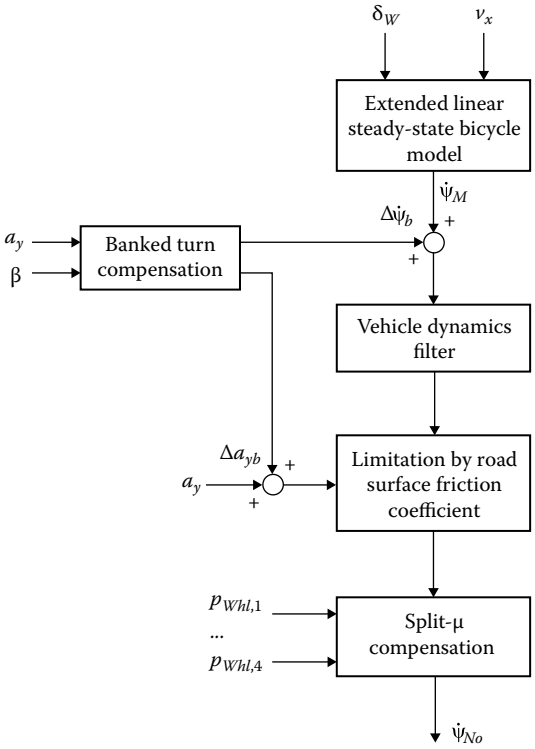


FIGURE 30.36 Derivation of the nominal yaw velocity of the vehicle.

The interventions are then also avoided. For violent driving maneuvers, however, the driver has still some support from the ESP system. See also [38] for an alternative analytic approach to determine the bank angle of the road.

Figure 30.36 shows the principle of the computation of the nominal yaw velocity.

The nominal yaw velocity is derived using the steady-state linear bicycle model, that is, for a coasting vehicle with a constant steering angle. During braking and lift-off of the accelerator pedal, the vehicle velocity is not constant and the vehicle tends to oversteer. Also, care must be taken here that ESP does not intervene before the physical limit is reached. Such an intervention is felt by the driver as a sudden deceleration of the vehicle. Also, the noise from the system may be noticed by him. Therefore, the nominal yaw velocity must be modified in these situations. A special situation is ABS control. Since during ABS control the vehicle decelerates already with maximum deceleration and the system already makes noise, ESP interventions are not noticed. In particular, slip increments that increase the slip hardly change the brake force and their influence can hardly be felt by the driver. Therefore, the nominal yaw velocity as computed by the bicycle model with its limitation by the coefficient of friction of the road surface (30.6) can be used for ABS without any problem.

30.1.5.3.2 Computation of the Vehicle Slip Angle

Since the slip angle of the vehicle is an important indicator of the vehicle stability, many proposals have been made for its estimation. Here, the proposal of [39] is discussed. Other proposals can be found in [40–43].

For the derivation of the vehicle slip angle, the signals from the sensors installed in the vehicle are used. These sensors measure the physical values in the body fixed coordinate system (X, Y) (Figure 30.37).

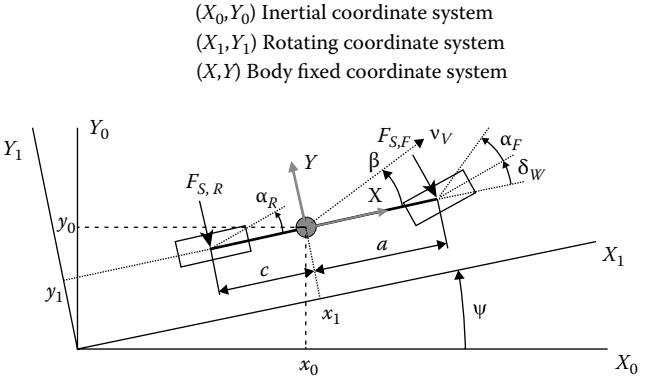


FIGURE 30.37 Definition of the inertial, reference, and body fixed coordinate systems.

The relations between the body fixed coordinate system and the inertial coordinate system (X_0, Y_0) are

$$\begin{aligned}
\dot{x}_0 &= v_x \cdot \cos\psi - v_y \cdot \sin\psi \\
\dot{y}_0 &= v_x \cdot \sin\psi + v_y \cdot \cos\psi \\
\ddot{x}_0 &= a_x \cdot \cos\psi - a_y \cdot \sin\psi \\
\ddot{y}_0 &= a_x \cdot \sin\psi + a_y \cdot \cos\psi
\end{aligned} \tag{30.7}$$

From Figure 30.37, it follows:

$$\begin{aligned}
&\tan(\beta+\psi) = \frac{\dot{y}_0}{\dot{x}_0}, \quad \beta = -\psi + \tan^{-1}\left(\frac{\dot{y}_0}{\dot{x}_0}\right) \\
&\dot{\beta} = -\dot{\psi} + \frac{1}{v_V{}^2} \cdot (\ddot{y}_0 \cdot \dot{x}_0 - \ddot{x}_0 \cdot \dot{y}_0) \\
&\dot{\beta} = -\dot{\psi} + \frac{1}{v_V}(a_y \cdot \cos\beta - a_x \cdot \sin\beta)
\end{aligned} \tag{30.8}$$

For small values of the slip angle, the equation for the slip angle velocity can be reduced and the slip angle obtained by direct integration.

$$\dot{\beta} = \frac{a_y}{v_V} - \dot{\psi}, \quad \beta(t) = \beta_0 + \int_0^t \left(\frac{a_y}{v_V} - \dot{\psi}\right) \cdot dt \tag{30.9}$$

Since the lateral acceleration and the yaw velocity are measured and the vehicle velocity is estimated, the vehicle slip angle can be computed if its initial value is known. However, since each of these signals may be biased, the computed vehicle slip angle error may grow in time to large values and the slip angle value must be frequently reset.

If the tire slips at the wheels are large, for example, during ABS, a Kalman filter may be used to obtain an alternative estimate of the vehicle slip angle [44]. For the Kalman filter, the dynamic equation for the lateral vehicle motion is used while the dynamic equation of the yaw motion of the vehicle is used to obtain the measurement equation. Figure 30.38 shows the idea of the Kalman filter. The estimates of the lateral velocity $\tilde{v}_y$ are obtained from the measurement equation. Since the sensor signals are corrupted by noise, they are filtered to obtain a filtered estimate of the lateral velocity $\hat{v}_y$. The filter equation is the transition function of the lateral velocity, which is the solution of the equation of lateral motion of the vehicle.

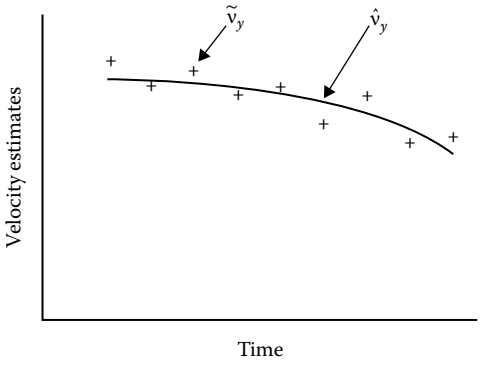


FIGURE 30.38 Principle of the Kalman filter for the vehicle lateral velocity with measurements $\tilde{v}_y$ and estimates $\hat{v}_y$.

The dynamic equations are based on a horizontal surface (no road slope) and wind forces are not considered (see Figure 30.37 for sign conventions).

$$m_V \cdot (\dot{v}_y + v_x \cdot \dot{\psi}) = -(F_{S1} + F_{S2}) \cdot \cos\delta_W - (F_{B1} + F_{B2}) \cdot \sin\delta_W - F_{S3} - F_{S4} \tag{30.10}$$

$$\begin{aligned} J_V \cdot \ddot{\psi} = &-(F_{S1} + F_{S2}) \cdot a \cdot \cos\delta_W - (F_{S1} - F_{S2}) \cdot b \cdot \sin\delta_W \\ &-(F_{B1} + F_{B2}) \cdot a \cdot \sin\delta_W + (F_{B1} - F_{B2}) \cdot b \cdot \cos\delta_W \\ &+(F_{S3} + F_{S4}) \cdot c + (F_{B3} - F_{B4}) \cdot b \end{aligned} \tag{30.11}$$

By using the friction ellipse, it can be shown that for small slip angles where $\tan\alpha \approx \alpha$ and for slip values that are not zero, a simple relation between the brake and side forces exists, which is also valid for large slip values, that is, for values in the nonlinear region of the μ-slip curve.

$$\begin{aligned} &F_B = \lambda \cdot \frac{F_R}{\sqrt{\lambda^2 + (C_\alpha / C_\lambda)^2 \cdot \alpha^2}}, \quad F_S = \alpha \cdot \left(\frac{C_\alpha}{C_\lambda}\right) \cdot \frac{F_R}{\sqrt{\lambda^2 + (C_\alpha / C_\lambda)^2 \cdot \alpha^2}} \\ &\Rightarrow F_S = \frac{C_\alpha \cdot \alpha}{C_\lambda \cdot \lambda} \cdot F_B \end{aligned} \tag{30.12}$$

Furthermore, by using the dynamic equation of the wheel velocity, the brake force can be related to the brake pressure at each wheel (the drive torque M_{CaHalf} has to be considered at the driven wheels only).

$$F_B = C_p \cdot \frac{p_{Whl}}{R} - \frac{M_{CaHalf}}{R} + \frac{J_{Whl}}{R^2} \cdot \dot{v}_{Whl} \tag{30.13}$$

If the slip angles at the front tires are small, then they are approximately equal. Similarly, if the slip angles at the rear tires are small, then they are also approximately equal. The slip angles can then be computed as

$$\alpha_F = \frac{v_y + a \cdot \dot{\psi}}{v_x}, \quad \alpha_R = \frac{v_y - c \cdot \dot{\psi}}{v_x} \tag{30.14}$$

Inserting (30.12–30.14) into (30.10 and 30.11) and after collecting terms, the equations can be written in reduced form by

$$\begin{aligned} \dot{v}_y &= A_{11} \cdot v_y + A_{12} \cdot \dot{\psi} + u_1 \\ \ddot{\psi} &= A_{21} \cdot v_y + A_{22} \cdot \dot{\psi} + u_2 \end{aligned} \tag{30.15}$$

These equations must be integrated to obtain the transition function and the measurement equation for the Kalman filter. Since the control function is realized as a discrete time controller, a numerical integration is a simple alternative to the analytic solution of the linear differential Equations 30.15. Here, the simplest method, the Euler method, is used.

$$v_{y,k+1} = (A_{11} \cdot T + 1) \cdot v_{y,k} + A_{12} \cdot T \cdot \dot{\psi}_k + u_{1,k} \cdot T \tag{30.16}$$

$$\dot{\psi}_{k+1} = A_{21} \cdot T \cdot v_{y,k} + (A_{22} \cdot T + 1) \cdot \dot{\psi}_k + u_{2,k} \cdot T \tag{30.17}$$

In order to use (30.17) as a measurement equation, a linear extrapolation of the yaw velocity is used.

$$\dot{\psi}_{k+1} = 2 \cdot \dot{\psi}_k - \dot{\psi}_{k-1} \tag{30.18}$$

If (30.18) is inserted into (30.17), then the measurement equation results

$$v_{y,k} = \frac{(1 - A_{22} \cdot T) \cdot \dot{\psi}_k - \dot{\psi}_{k-1} - u_{2,k} \cdot T}{A_{21} \cdot T} \tag{30.19}$$

Figure 30.39a shows a comparison between the estimated slip angle at the front wheels and the measured value, while Figure 30.39b shows the comparison at the rear wheels for an ABS braking and double-lane change maneuver on ice. The comparison shows an excellent result of the estimation. Also, for large slip angles, the estimation is close to the measured values. On ice, the μ-slip angle curve has its maximum at approximately 5°. This shows that the estimation produces good results also for slip angles which are in the nonlinear and in the unstable region of the μ-slip angle curve. Since in all equations the quotient of the lateral and the longitudinal tire stiffness is used (30.12), the estimation of the slip angle is robust with respect to changes in the tire properties. For instance, if winter tires are used instead of summer tires, then both the longitudinal and the lateral tire stiffness decrease significantly as compared with those of summer tires; the quotient,

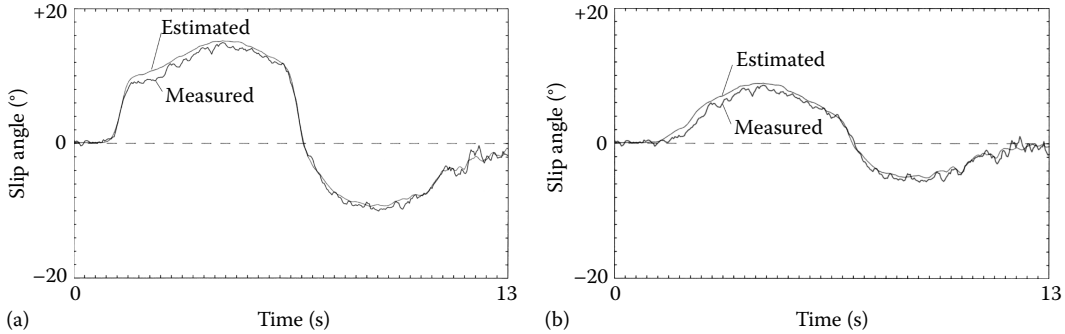


FIGURE 30.39 Comparisons between measured and Kalman-filtered estimates of the slip angles of the car at the front axle (a) and at the rear axle (b) during an ABS braking with double-lane change maneuver on ice.

however, does not change significantly. Similarly, if the tires are worn, the longitudinal and lateral stiffnesses increase significantly as compared with those of new tires; however, the change in the quotient is not significant. It is therefore not required to change the estimation depending on the tire make and state.

30.1.5.3.3 Control Algorithm

A PID control is used to compute the required yaw moment. Inputs to the PID controller are the actuating signals of the slip angle and the yaw velocity. In situations where the slip angle cannot be computed with sufficient accuracy, the control of the yaw velocity becomes dominant. As has been shown earlier, it is difficult to obtain the nominal value of the yaw velocity. Moreover, the vehicle handling behavior and therefore also the value of the yaw velocity may change significantly with changing lateral tire stiffnesses (winter/summer tires) and with changing vehicle mass and changing location of the center of mass. These changes are unknown to the controller and thus interventions may occur within the physical limit because of the different basic handling behavior of the vehicle. But these early interventions must be avoided. Often, such early interventions are avoided by enlarging the dead zone for the yaw velocity.

The yaw moment is generated by slip changes at selected wheels. If the slip values are well below the value λ_0, the nominal slip value is converted to a nominal wheel torque value and the wheel torque is controlled to this nominal value. The required slip changes at the wheels are chosen such that the sum of the yaw moment changes from the selected wheels equals the required yaw moment on the vehicle. The slip change at a wheel is proportional to the gain between the yaw moment and the slip change. The gain may be computed from the friction ellipse (30.12) using the actual slip and slip angle values.

During braking and driving on a split-μ road surface, a yaw moment on the vehicle is generated because of the difference in longitudinal forces on the wheels on the left and on the right side of the vehicle. This yaw moment is directly controlled by the vehicle dynamics controller by specification of the brake torque (or pressure) difference between the left and the right side of each axle. For braking maneuvers, the time derivative of the brake torque difference is also specified (Figure 30.40). In Figure 30.40, the maximum pressure differences Δp_1 and Δp_2 and the time points t_1 and t_2 as measured from the start of the braking maneuver can be chosen freely for each axle to fulfill the requirements (the braking maneuver starts at time point $t = 0$ s). For driving maneuvers, the vehicle dynamics controller reduces the brake torque difference by ΔM_{NoLock} from that set as target by the TSC M_{NoLock}.

Instead of controlling the traction slip individually of each driven wheel, the TSC controls the average slip λ_{MA} and the difference slip of the driven wheels of the driven axle. If the vehicle motion differs from the nominal motion, the vehicle dynamics controller reduces the average slip

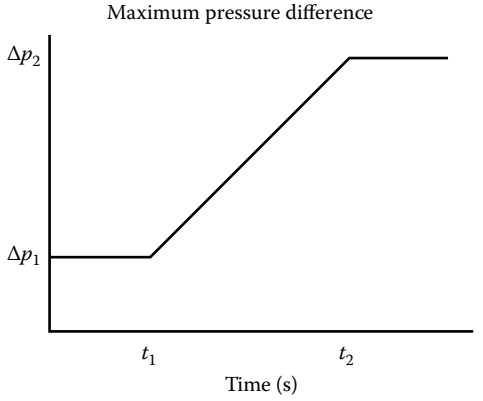


FIGURE 30.40 Limitation of the brake pressure difference between the wheels at one axle for split-μ braking.

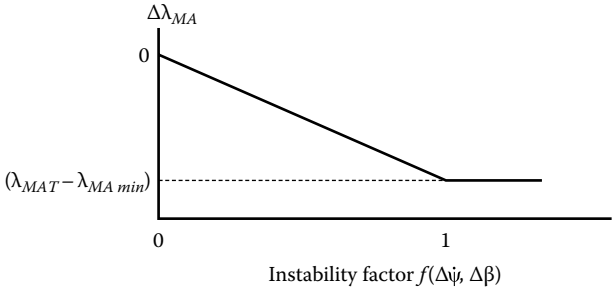


FIGURE 30.41 Reduction of the cardan slip from its target value depending on the stability margin of the vehicle.

by sending a slip increment (reduction) to the TSC $\Delta\lambda_{MA}$. This will then reduce the engine output torque. The reduction depends on an instability factor $f(\Delta\beta,\Delta\dot{\psi})$ which itself depends on the stability margin of the vehicle (Figure 30.41). The instability factor also considers the limit in the engine torque (its maximum value, idle torque, idle rotational velocity). The larger the instability factor f, the smaller the stability margin. If the instability factor is 0, then the average traction slip is not reduced from its target value, which is the value for optimal traction and stability λ_{MAT}. If the instability factor is 1, then the slip increment reduces the average slip from the target slip to the velocity-dependent minimal value for maximum vehicle stability λ_{MAmin}. Further details will be discussed in Section 30.1.5.5.

For each vehicle situation, wheels that are selected to generate the yaw moment are defined. There is some ambiguity during the transition from one vehicle situation to another one, for example, from coasting to partial or full braking. In each vehicle situation, different wheels are selected to generate the yaw moment. The transitions must be carefully tested and adjusted so as to not switch too early or too late from one selection to the other. Since the controllers do not change their structure with the driving situation and since in all controllers the slip is the fundamental variable, the hierarchical control structure presented seems to be the right choice to cope with these transition problems best.

30.1.5.4 Brake Slip Controller

The BSC is used to control the wheel slip for positive values (brake slip). The controller consists of several main tasks (Figure 30.42). These are as follows:

- Estimation of the vehicle velocity.
- Determination of the target wheel slips for ABS.
- Control of the wheel slip.
- Determination of the valve stimulations.

In Figure 30.42, the input and output variables are shown. The slip increment $\Delta\lambda$ is supplied for each wheel by the vehicle dynamics controller to control the yaw moment on the vehicle.

30.1.5.4.1 Estimation of the Vehicle Velocity

For the determination of the actual value of the wheel slip, the free rolling wheel velocity v_{WhlFre} is required. The idea of calculation of the free rolling wheel velocity is given as follows. If during certain time periods in the slip control the brake torque is kept so low that a stable slip results, then by extrapolation of the brake torque to a value which is zero the free rolling wheel velocity can be evaluated. The time periods are called "adaptation phases." From the free rolling wheel velocity, the longitudinal velocity of the center of mass of the vehicle v_x can be computed using the lateral and the yaw velocity of the vehicle and the steering angle of the wheel. From the velocity of the vehicle center of mass, the free rolling velocities of the other wheels can be subsequently computed.

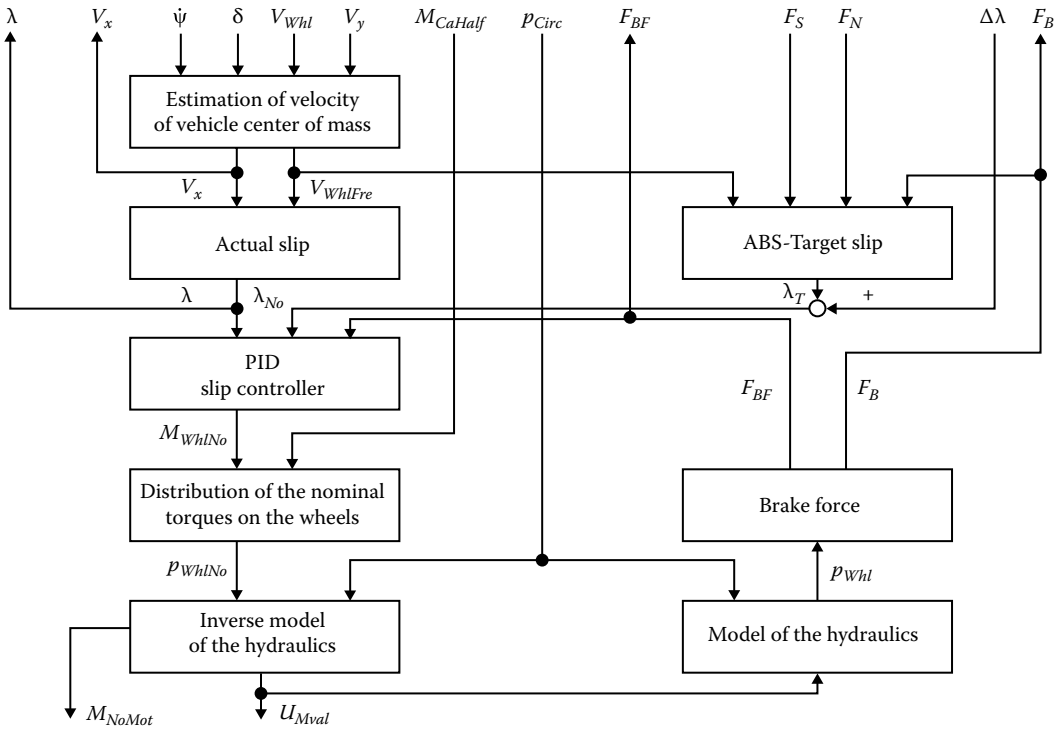


FIGURE 30.42 Block diagram of the BSC.

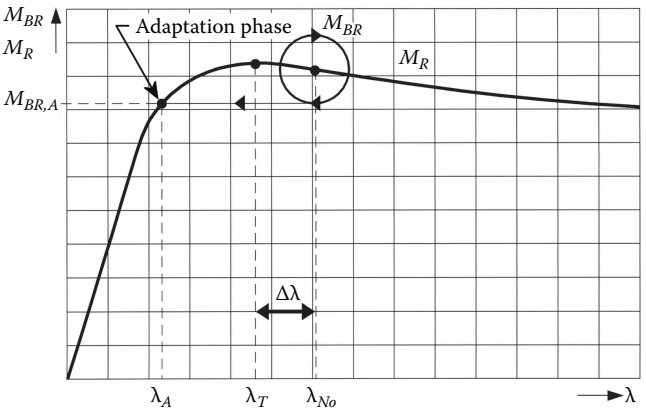


FIGURE 30.43 Determination of the free rolling wheel velocity during the adaptation phase of an ABS-braking maneuver.

The idea is shown in Figure 30.43. The circle symbolizes the brake torque modulations in order to maintain the wheel slip λ close to the nominal slip value λ_{No}. During the adaptation phase, the brake torque is kept fixed at its lowest value $M_{BR,A}$ during the modulation. The stable slip λ_A results. If the brake torque $M_{BR,A}$ is known, then the brake force $F_{B,A}$ during the adaptation phase can be computed. In order to eliminate influences from variations in the normal force F_N during the adaptation phase, the computation uses the μ-slip curve and the coefficient of friction μ_A instead of the brake force–slip curve. Furthermore, it is assumed that the point (λ_A, μ_A) lies on that portion of the μ-slip curve where it can be approximated by a straight line from the origin (0,0).

$$\mu_A = \frac{F_{B,A}}{F_N} = C_\lambda \cdot \lambda_A = C_\lambda \cdot \frac{v_{WhlFre} - v_{Whl,A}}{v_{WhlFre}}$$
$$\Rightarrow v_{WhlFre} = v_{Whl,A} \cdot \frac{C_\lambda}{C_\lambda - (F_{B,A} / F_N)} \tag{30.20}$$

At every sample point in time, the free rolling wheel velocity and the vehicle longitudinal velocity v_x can thus be estimated. Usually the left and the right rear wheels are alternatively in the adaptation phase. This improves the vehicle stability and the loss in braking distance is not as large as if one of the front wheels is used for the adaptation phase. For best estimation results, there should always be one of the wheels in the adaptation phase. If that is not possible, then between the adaptation phases the vehicle velocity is obtained by linear extrapolation. Figure 30.44b shows the pressures at the rear wheels and the rear wheel velocities with the adaptation phases during a straight-line ABS-braking maneuver from 120 km/h on a dry and smooth asphalt road surface.

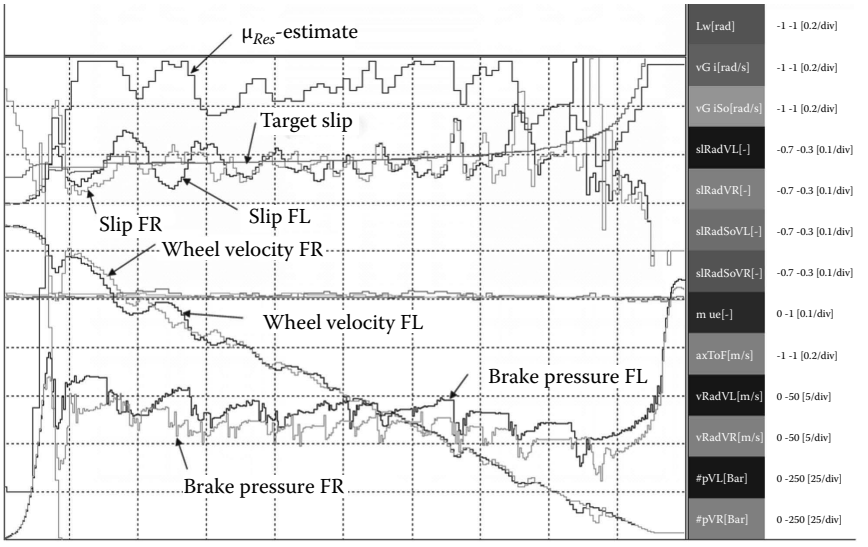
Since the above estimates of the vehicle velocity are error prone, a Kalman filter is used to filter the estimates. For the filter equation, the equation of motion of v_x is used.

$$\dot{v}_x = \frac{1}{m_V} \cdot \left\{ (F_{S1} + F_{S2}) \cdot \sin\delta_W - (F_{B1} + F_{B2}) \cdot \cos\delta_W - (F_{B3} + F_{B4}) \right\}$$
$$- \frac{c_w \cdot A \cdot v_x^{\,2} \cdot \rho / 2}{m_V} - \dot{v}_{x,off} \tag{30.21}$$
$$\ddot{v}_{x,off} = 0$$

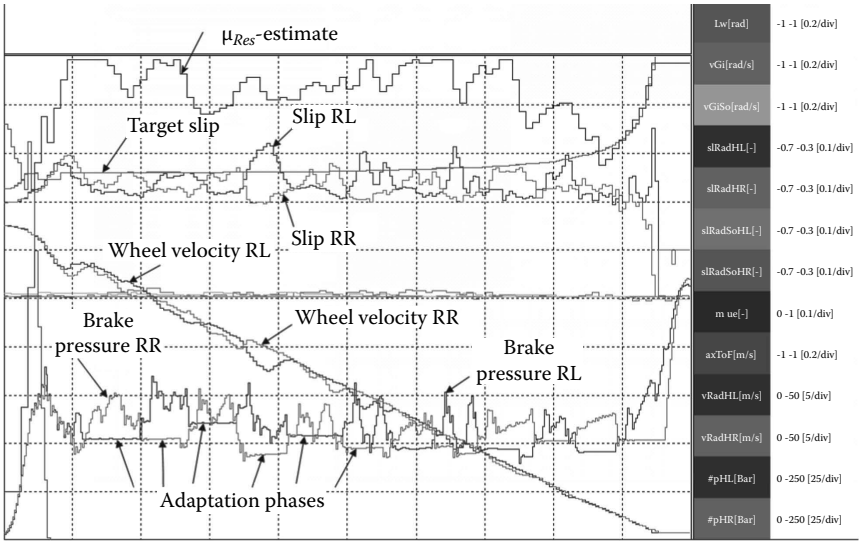
In the acceleration term, the product of the lateral velocity and the yaw velocity is neglected as compared with the time rate of change of the longitudinal velocity. The slope of the road, $\dot{v}_{x,off}$, is assumed to be constant and is also estimated by the Kalman filter. The brake and the side forces are estimated by (30.12) and (30.13).

At this point, the evaluation of the vehicle longitudinal velocity v_x will be discussed for the other driving situations also. In these other driving situations, one or more wheels may have a slip value which is in the stable portion of the μ-slip curve, such as during partial braking. Such wheels may then be used to determine the free rolling wheel velocity by extrapolation immediately, that is, without reduction of the brake torque. If there is more than one such wheel, then more than one evaluation of v_x may be obtained at each sample time point. The selection of the best estimate of v_x out of the two, three, or four estimates of v_x has been made on the basis of a selection logic. This selection logic is considered based on the following: if the brake or the accelerator pedal is applied, if the switch "ESP off" has been activated, the magnitude of the engine torque and the engine rotational velocity, the magnitude of the brake pressure at the wheel, stable or unstable behavior of the wheel, the magnitude of the sensor signals, and so on. Since the selection logic is very complex, in particular for four-wheel drive vehicles, it has been replaced by fuzzy logic.

The vehicle velocity estimation using fuzzy logic takes the v_x estimate of each wheel and computes a weighted average. The weight of each estimated v_x is determined by fuzzy methods and depends of the stability probability of each wheel. The weight is large if the wheel behavior is very stable and small if the wheel behavior is nearly unstable. The stability probability is judged on the basis of the wheel velocity, the wheel acceleration, and on the basis of the wheel jerk. The weighted average is then used and filtered by the Kalman filter.



(a)



(b)

FIGURE 30.44 Time histories of some braking variables at the front axle (a) and the rear axle (b) showing the adaptation phases at the rear wheels during a straight-line ABS-braking maneuver.

30.1.5.4.2 Determination of the Target Wheel Slips for ABS

As the first approach, the target brake slip for each wheel is determined by the BSC to achieve minimal stopping distances during full braking. Since the brake slip at which the brake force is at its maximum is larger for high friction road surfaces than for low friction road surfaces, the target slip is computed depending on the coefficient of the road surface. Figure 30.45 shows the idea. In this figure, two μ-slip curves are drawn: one for a low friction and one for a high friction road surface, each with their own maximum. The idea is that the line which connects the two maxima indicates at which slip value the maximum of the μ-slip curve of all other road surfaces lies. So, if the coefficient of friction of the road surface is μ_{Res}, then the slip value at the maximum of the μ-slip curve of the road surface is said to be $\lambda_0(\mu_{Res})$.

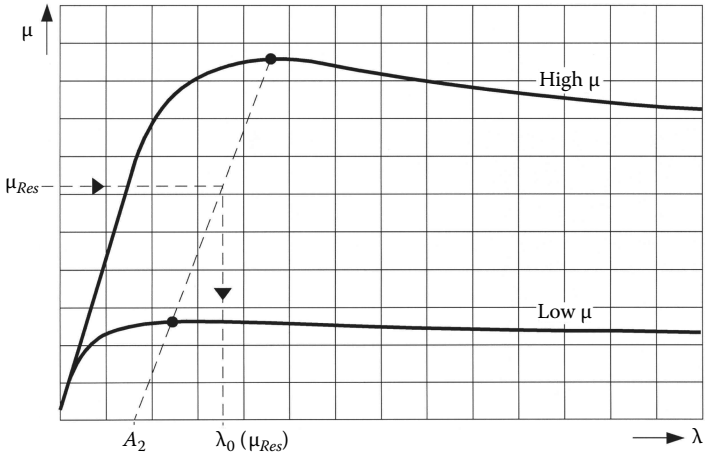


FIGURE 30.45 Computation of the target slips of the wheels for an ABS-braking maneuver.

For low, free rolling wheel velocities, the target slip is increased from λ_0 to improve the slip control. For improved stability of the vehicle, the target slip of the wheels at the rear axle is taken to be smaller than that for the wheels at the front axle.

$$\lambda_T(\mu_{Res}) = A_2 + A_0 \cdot \mu_{Res} + \frac{A_1}{v_{WhlFre}}$$
$$\mu_{Res} = \frac{\sqrt{F_B{}^2 + F_S{}^2}}{F_N} \tag{30.22}$$

Figure 30.44 shows the estimate of the coefficient of friction of the road surface μ_{Res} and the corresponding determination of the target slip of the wheels on the front and on the rear axles. It can be seen from this experiment that the target slip is found very quickly from the beginning of the braking maneuver.

30.1.5.4.3 Control of the Wheel Slip

The nominal wheel slip λ_{No} results from the addition of the target slip $\lambda_T(\mu_{Res})$ in case of ABS and the slip increment $\Delta\lambda$ requested by the vehicle dynamics controller. If the steady-state brake force at the nominal wheel slip λ_{No} is denoted by F_{BF}, then the nominal brake torque at the wheel is computed by (30.23) in which "SUM" stands for the time integral.

$$M_{WhlNo} = F_{BF} \cdot R + K_p \cdot (\lambda_{No} - \lambda) \cdot R + K_d \cdot (\dot{v}_{Whl} - \dot{v}_{WhlFre}) \cdot \frac{J_{Whl}}{R}$$
$$+K_i \cdot C_p \cdot SUM\left\{(\lambda_{No} - \lambda) \cdot D_T\right\} \tag{30.23}$$

The steady-state brake force at the nominal wheel slip is estimated by low pass filtering of the brake force, which is computed by (30.13).

$$T_1 \cdot \dot{F}_{BF} + F_{BF} = F_B \tag{30.24}$$

From the nominal brake torque, the nominal brake pressure can readily be found by consideration of the drive torque on the wheel M_{CaHalf}, if the wheel is a driven wheel.

$$p_{WhlNo} = \frac{M_{WhlNo} + M_{CaHalf}}{C_p} \tag{30.25}$$

If this brake pressure is negative, then the drag torque of the engine on the driven wheel is too high and the engine output torque must be increased to M_{NoMot}. In the evaluation of the nominal engine torque, a dynamic term for the acceleration of the power and drive train is included. For the computation of the engine torque, the driven wheel with the largest absolute value of the drag torque is considered (i.e., with the minimum drive torque).

$$M_{NoMot} = -\frac{2 \cdot m}{I_{Tr}} + \frac{J_{Mot} \cdot I_{Tr}}{R} \cdot \dot{v}_x \tag{30.26}$$

$$m = \text{MIN}(M_{Whl3No}, M_{Whl4No}) \quad \text{(for a rear wheel driven vehicle)} \tag{30.27}$$

For a good slip control performance, it is important that the steady-state brake force F_{BF} is found as quickly as possible. Otherwise, the control may start to oscillate which has a negative influence on the control performance (stopping distance, comfort).

30.1.5.4.4 Determination of the Valve Stimulations

If the nominal wheel brake pressure is positive, then by using a model of the hydraulic system the valve stimulation time periods can be obtained. The simplest model is based on Bernoulli's law for the flow of incompressible fluids. In order to compute the times during which an inlet or AV must be opened, the inverse model is used.

$$U_{Mval} = \frac{p_{WhlNo} - p_{Whl}}{(x_1 + x_2 \cdot p_{Whl}) \cdot \sqrt{|p_{Circ} - p_{Whl}|}}$$

$$U_{Mval} > 0 \quad \text{Pressure increase} \tag{30.28}$$

$$U_{Mval} = 0 \quad \text{Pressure hold}$$

$$U_{Mval} < 0 \quad \text{Pressure decrease}$$

The parameters x_1 and x_2 depend on the volume–pressure characteristic curve and on the throttle bores of the hydraulic system (valves). The computed positive value U_{Mval} is the time period the EV is opened. If the value U_{Mval} is negative, then its absolute value is the time period the AV is opened. Since the valve current periods can only be realized for certain discrete values (1, 2, 3 ms, …), the value U_{Mval} is rounded and the resulting actual pressure value at the wheel brake p_{Whl}, because of the rounded stimulation time periods, must be computed. For the computation, the model of the hydraulic system is used again. Using the brake pressure value, the brake force F_B can be computed by (30.13).

30.1.5.5 Traction Slip Controller

The TSC controls the negative slip (drive slip) of the driven wheels. Instead of controlling the wheel drive slip for each tire individually, the TSC controls the average slip of the driven wheels λ_{MA} and the dead zone for the difference in the drive slip of the driven wheels D_λ. The idea behind this choice is that the average slip can be controlled by the engine output torque and the slip difference by the brake torque at the fastest wheel. Thus, the control is split into a slow part (engine control with the large inertia of the power and drive train) and a fast part (brake control with a small inertia of one

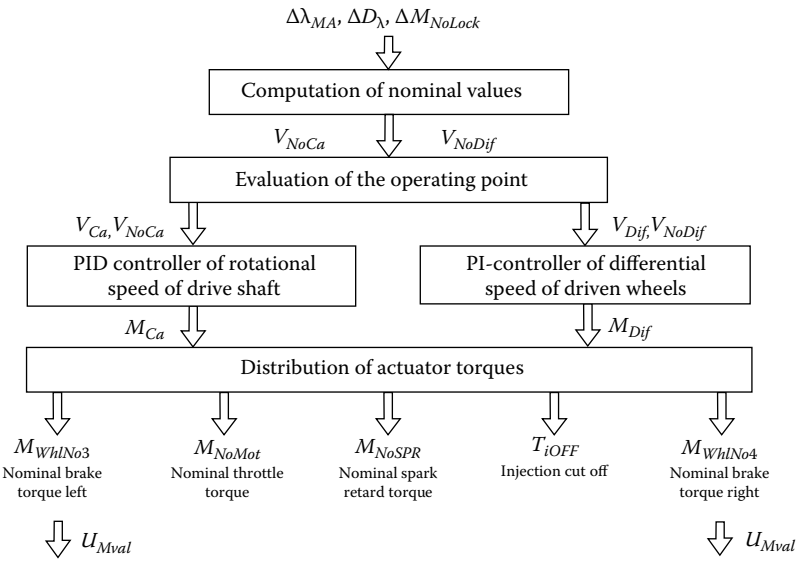


FIGURE 30.46 Block diagram of the TSC.

wheel of the driven axle). The principle of the TSC will be explained for a rear wheel driven car first. The TSC consists of two main parts (Figure 30.46):

- The average slip controller (also called the cardan slip controller), with output, the nominal engine output torque.
- The difference slip controller, with output, the nominal brake torque at one wheel.

30.1.5.5.1 Average Slip Controller

This part of the TSC controls the average value of the slips of the driven wheels by controlling the engine output torque. The target value of the average slip λ_{MAT} is computed similarly as for ABS by (30.22). Similar to the ABS function of the BSC, the TCS function is started if the average slip of the driven wheels exceeds the target slip. The nominal value λ_{MA} of the average slip is computed by adding the increment value $\Delta\lambda_{MA}$ computed by the vehicle dynamics controller to the target value λ_{MAT}. Using the average value of the free rolling velocities of the driven wheels and dividing it by $(1 + \lambda_{MA})$, the nominal average velocity of the driven wheels v_{NoCa} is computed (30.30). The actuating signal for the cardan slip controller is then $v_{NoCa} - v_{Ca}$ in which v_{Ca} is the average velocity of both driven wheels.

$$v_{Ca} = \frac{v_{Whl,3} + v_{Whl,4}}{2} \quad \text{(for a rear wheel drive vehicle)} \tag{30.29}$$

The actuating signal is the input of a PID controller. The output of the PID controller is the torque of the cardan axle multiplied by the ratio of the drive train M_{Ca} from which the engine torque can be easily computed by using the total transmission ratio of the power and drive train I_{Tr}.

The gains of the PID controller depend on the following driving situations:

- The gear position of the transmission
- The efficiency of the converter if the vehicle has an automatic transmission
- The distance of the actual engine torque value to its limits (maximum/minimum values)

- The gradient with which the hydraulic system can generate active brake pressure at the driven wheels
- The coefficient of friction of the road surface
- The need for side forces on the driven wheels

Since the drive train can oscillate with a damping factor whose value can be very low, the gains of the PID controller must be chosen to be small enough in order to avoid control instability. Furthermore, the average velocity of the driven wheels is filtered by a low pass filter in order to make the controller insensitive to these oscillations. The frequency of the oscillations depends on the gear position of the transmission and can be as low as 2.0 Hz. Therefore, the bandwidth of the low pass filter is chosen to be 2 Hz.

Because of the low gains of the controller and because of the low pass filtering of the average wheel velocities, changes in the cardan slip cannot be realized fast enough for the required TCS vehicle stability and vehicle dynamics control performance. The reduction in the engine output torque may be so slow that the wheel with the larger traction force will also start to spin. Therefore, the engine output torque is also controlled by a fast parallel control depending on the behavior of the wheel with the larger traction force, that is, if the wheel starts to spin. The smaller one of the two torques of the cardan axle (output of the PID controller and fast reduction of the cardan torque) is passed on as M_{Ca}.

The reduction in the drive torque at the wheels may be realized by reducing the throttle angle, by reducing spark advance, by reducing the amount of fuel injected, and sometimes by fuel injection cut off. Because of the legal requirements on the limitation of the engine exhaust emissions, modification of the spark timing and the amount of fuel injected are not always allowed. The reduction in the drive torque on the driven wheels may then not be fast enough. In that case, both driven wheels may be actively braked in order to reduce the net drive torque on them. This brake intervention to support the decrease in the drive torque on the driven wheels is called the "symmetric brake intervention." The required engine torque is sent to the engine controller where the interventions in the throttle angle, the spark timing, and the injection amount are controlled in an integrated manner. The actual engine output torque is computed by the engine controller and its value is sent to the ESP-control unit by the controller area network (CAN) bus.

30.1.5.5.2 Difference Slip Controller

The other part of the TSC controls the slip difference between the driven wheels to be smaller than D_λ by controlling the brake torque on the wheel with the lowest traction force. This control maximizes the traction forces on the wheels during cornering and on split-μ road surfaces. Although for straight-line driving on a homogeneous road surface D_λ can be chosen to be zero for optimal traction for driving in a turn or on split-μ road surfaces, this choice cannot be made. Because the normal force on the wheel on the inside of the turn is smaller than that on the wheel on the outside of the turn, the slip value of the wheel on the inside of the turn will be larger than that of the wheel on the outside of the turn, although the drive torques on both wheels are equal. A similar situation exists during driving on a split-μ road surface. In both situations, there exists always a difference in slips between the left and the right wheels but interventions must not occur if no wheel spins. Only if the difference in the slips between both wheels exceeds a threshold (also called dead zone) D_λ, the slip difference control is activated to reduce the slip difference to a value below this threshold. The threshold D_λ specifies a dead zone within which no brake control happens. Instead of controlling the difference in the slips between both wheels, the difference in the wheel velocities is controlled. During straight-line driving, D_λ is zero and the difference in the nominal velocities between the two wheels v_{NoDif} is also zero.

During cornering, the situation is more complex because the wheel on the inside of the turn has a smaller free rolling velocity than the wheel on the outside of the turn. If wheel 3 is on the outside and wheel 4 on the inside of the turn, then v_{NoDif} can be evaluated as follows.

$$\begin{aligned}
\lambda_{WhlNo3} &= \lambda_{MA} = \frac{v_{WhlFre3} - v_{WhlNo3}}{v_{WhlNo3}} \\
\lambda_{WhlNo4} &= \lambda_{MA} + D_{\lambda} = \frac{v_{WhlFre4} - v_{WhlNo4}}{v_{WhlNo4}} \\
v_{WhlNo3} &= \frac{v_{WhlFre3}}{1+\lambda_{MA}} \\
v_{WhlNo4} &= \frac{v_{WhlFre4}}{1+\lambda_{MA}+D_{\lambda}} \\
v_{NoDif} &= v_{WhlNo4} - v_{WhlNo3} \\
v_{Dif} &= v_{Whl,4} - v_{Whl,3} \\
e_{\Delta v} &= v_{Dif} - v_{NoDif} \quad \text{(actuating signal)}
\end{aligned} \tag{30.30}$$

If the actuating signal is negative, that is, if the difference in the wheel velocities is smaller than the nominal difference, then v_{Dif} lies inside the dead zone and the controller is not activated. Otherwise the velocity difference is controlled with a resulting brake torque M_{NoLock} which works like a differential lock. From the brake torque, the valve stimulations can be computed by using the model of the hydraulic system described at the BSC.

If the brake at one of the driven wheels is activated, then a yaw moment on the vehicle results. If the yaw moment on the vehicle must be reduced, the vehicle dynamics controller first reduces this yaw moment since this reduction is more comfortable than a brake intervention at a front wheel. The reduction from the vehicle dynamics controller is ΔM_{NoLock}. This reduction will increase the slip difference between the driven wheels. Therefore, the vehicle dynamics controller must also increase the dead zone by ΔD_{λ} at the same time when the vehicle dynamics controller requests a reduction in the brake torque. Otherwise the slip difference controller will immediately increase the brake torque again.

30.1.5.5.3 Four-Wheel Drive Vehicles

The principle of the TSC for two-wheel drive vehicles can be extended to result in the principle of the controller for four-wheel drive vehicles. Figure 30.47 shows the principle of a typical four-wheel drive train. After the transmission 2, the central cardan torque is split by the central differential 3 between the front cardan axle and the rear cardan axle. The front cardan axle torque is split by the differential 4 between the left and the right wheels at the front axle while the rear cardan axle torque is split by the differential 5 between the left and the right wheels at the rear axle. The differential 3 or 5 may have a coupling like a viscous coupling in order to connect the rear axle with the power train.

Figure 30.48 shows the principle of the TSC. The longitudinal lock of the brake controller at the front axle can reduce the difference velocity between the front cardan axle and the rear cardan axle if the front cardan axle spins by symmetric braking of the front wheels. Similarly, the longitudinal lock of the brake controller at the rear axle can reduce the difference velocity between the front cardan axle and the rear cardan axle if the rear cardan axle spins by symmetric braking of the rear wheels. By symmetric braking of all four wheels, the engine output torque (or cardan torque) can be reduced. The lateral lock function of the brake controller reduces the velocity difference between the wheels at the front and the rear axle.

Figure 30.49 shows the control of a four-wheel drive vehicle during drive-off on a split-μ uphill road surface with a slope of 20°. The engine torque request from the driver is seen to be a steady increasing torque M_{Dr}. After about 1.2 s, the front left wheel on the slippery road surface suddenly starts to spin (Figure 30.49a). Both the brake at this wheel is actuated $p_{Whl,1}$ and the torque at the

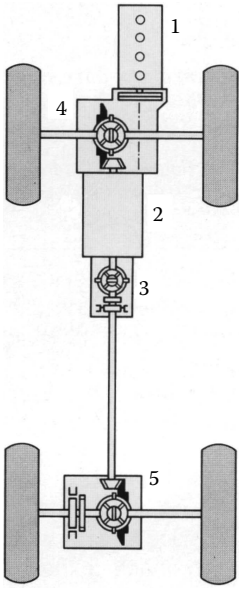


FIGURE 30.47 Principle of a four-wheel drive and power train with optional viscous clutches at the central and rear axle differentials (1) engine, (2) transmission, (3) central differential, (4) differential at the front axle, (5) differential at the rear axle.

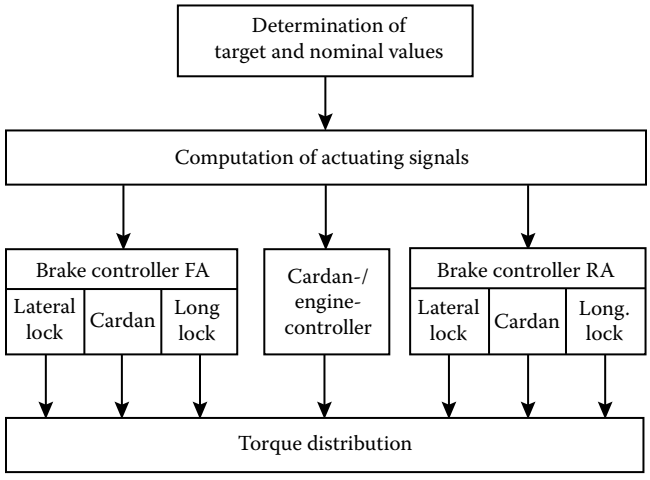
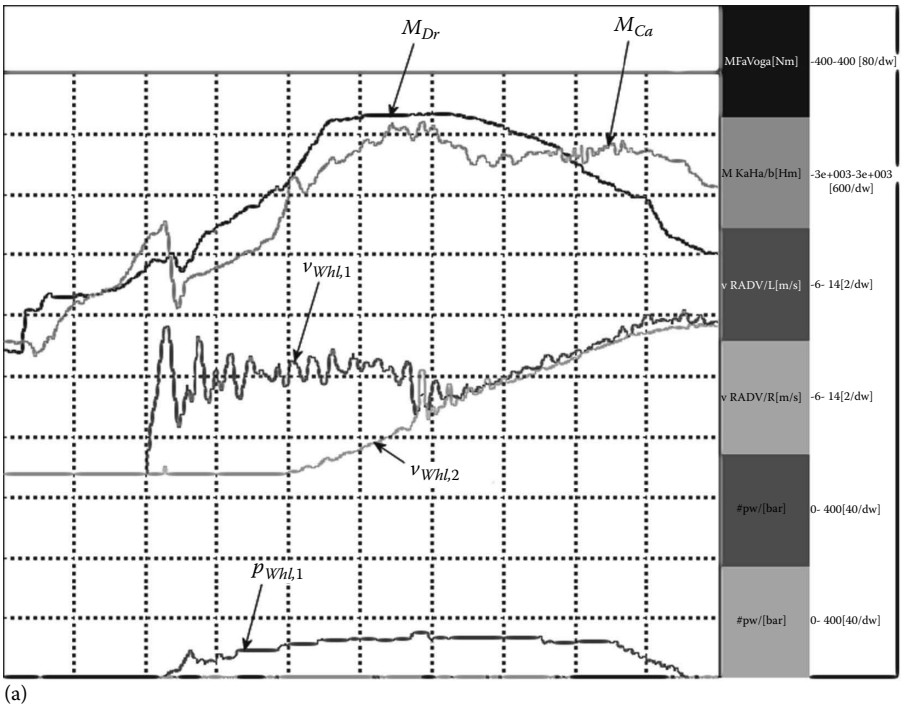
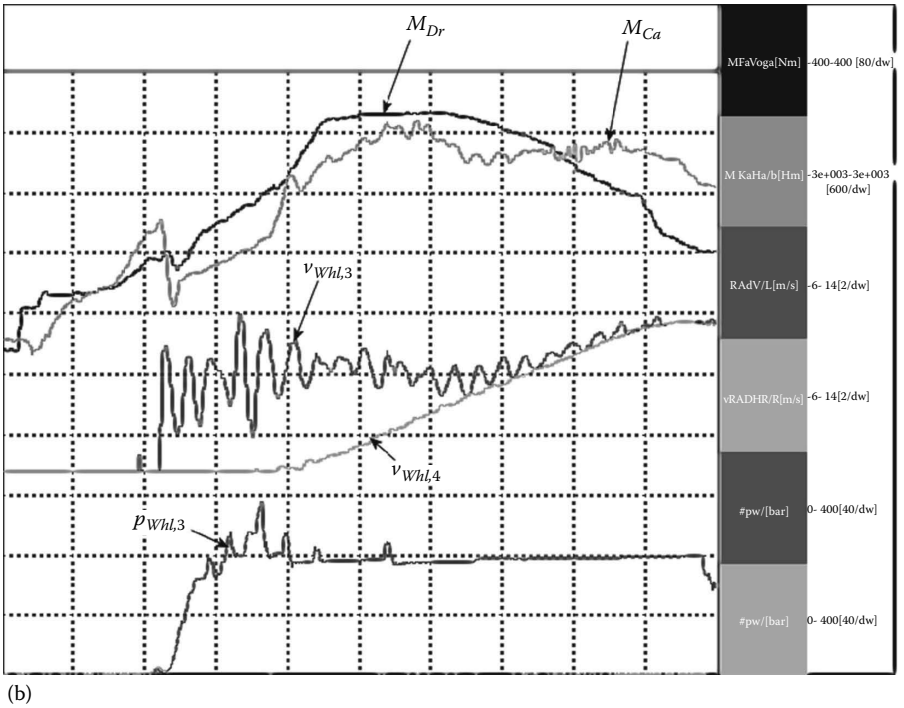


FIGURE 30.48 Block diagram of the TSC for a four-wheel drive vehicle.

central cardan axle M_{Ca} is reduced by the control system. The delay in the control is clearly visible. Because of the strong filtering of the wheel velocity, the pressure at the wheel does not follow the oscillations in the wheel velocity. A similar situation can be found at the rear axle (Figure 30.49b). As the brake torques at the left wheels increase, the torque at the central cardan axle can be increased again. The right front wheel starts to rotate after about 4.8 s and so does the right rear wheel, that is, the wheel velocities are then larger than the minimum velocity that can be measured (2.75 km/h). It is likely that at this time the vehicle already moves. After about 7 s, the difference in the velocity between the two front wheels has dropped to a small value and the wheel brake pressure at the front wheel can be kept constant. At the end of the maneuver, the driver releases the accelerator pedal and the brake pressure at the left wheels can be reduced. Because the rear brakes are smaller than the front brakes, the pressure at the rear brake is higher than that at the front brake.



(a)



(b)

FIGURE 30.49 Time histories of some variables of a four-wheel drive vehicle on a split-μ road surface during an uphill drive-off maneuver: (a) front axle, and (b) rear axle.

30.1.5.6 Results

Figure 30.50 shows the results from an ABS maneuver with a rear wheel drive test vehicle. The results are obtained from a closed-loop test maneuver in which the vehicle performs a lane change during full braking on polished ice ($\mu \cong 0.15$). The maneuver starts with an initial speed of 50 km/h, and braking and steering start at the same point in time. The test vehicle was alternatively equipped with series ABS and ESP. In Figure 30.50, the results of the maneuver using ABS are depicted on the left while those using ESP are shown on the right. The results of the maneuver with ABS show that both the yaw velocity and the vehicle slip angle attain large values right from the beginning so that the driver after 4 s has to quickly and heavily countersteer. He is barely able to bring the vehicle to a stable stop on the other lane. The results of the maneuver with ESP shown on the right of Figure 30.50 prove the superior handling performance of the vehicle. The driver need not countersteer nor does the steering angle have to be large, so the driver has an easy job. Also, both the yaw velocity and the vehicle slip angle follow the driver's steering input immediately and stay within a small range of values. The vehicle performs almost like it would do if the coefficient of friction would be much larger and therefore the driver is not surprised by a sudden change in the vehicle behavior. Surprisingly enough though is that the stopping distance of the ESP vehicle with the considerably improved handling performance is also considerably shorter than that of the more difficult to handle ABS vehicle. This can be explained by the limitation of the vehicle slip angle by the ESP system to small values and by the improved ABS regulation.

Figure 30.51 shows the simulation results of a rear wheel drive vehicle running on a circular track while the vehicle speed is increased slowly. Comparison is made between a conventional vehicle without ESP (Figure 30.51a) and the same vehicle with ESP (Figure 30.51b). The track is homogeneous and the coefficient of friction is high ($\mu = 1.0$). This is a closed-loop maneuver in which the driver has to keep the vehicle on the track. The diagrams show the required steering wheel angle, the resulting vehicle slip angle, and the resulting lateral deviation from the track. The dashed curves in both graphs are the collections of stationary points which are iteratively computed for each vehicle speed by the simulation program through variations of the steering wheel angle and the throttle valve angle. These dashed curves represent limit curves for the maneuver with the slowly increasing speed.

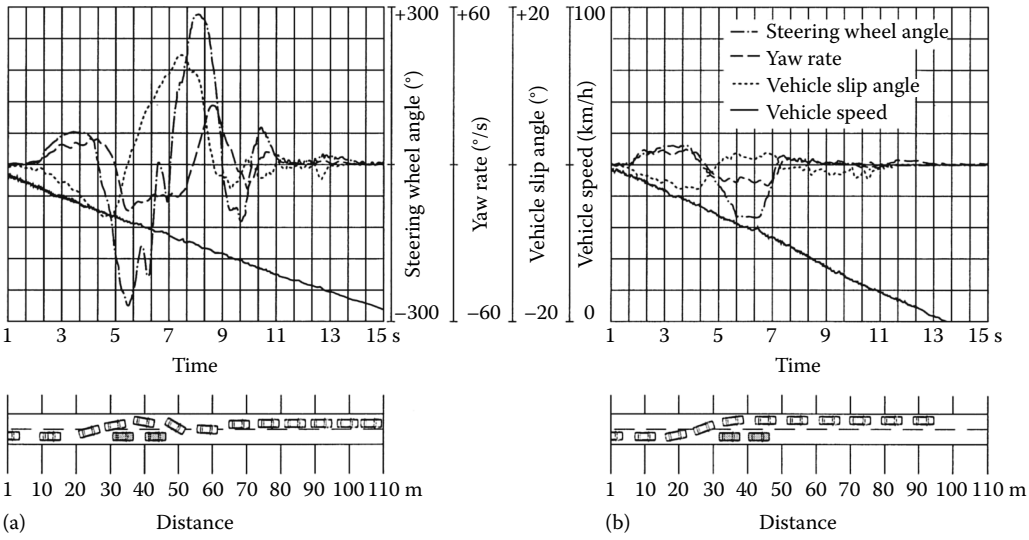


FIGURE 30.50 Comparison of the performance between stand-alone ABS and ESP during an ABS braking maneuver including a lane change. (a) Lane change with ABS ($V_V \approx 50$ km/h, $\mu \approx 0.15$) Full braking. (b) Lane change with ESP ($V_V \approx 50$ km/h, $\mu \approx 0.15$) Full braking.

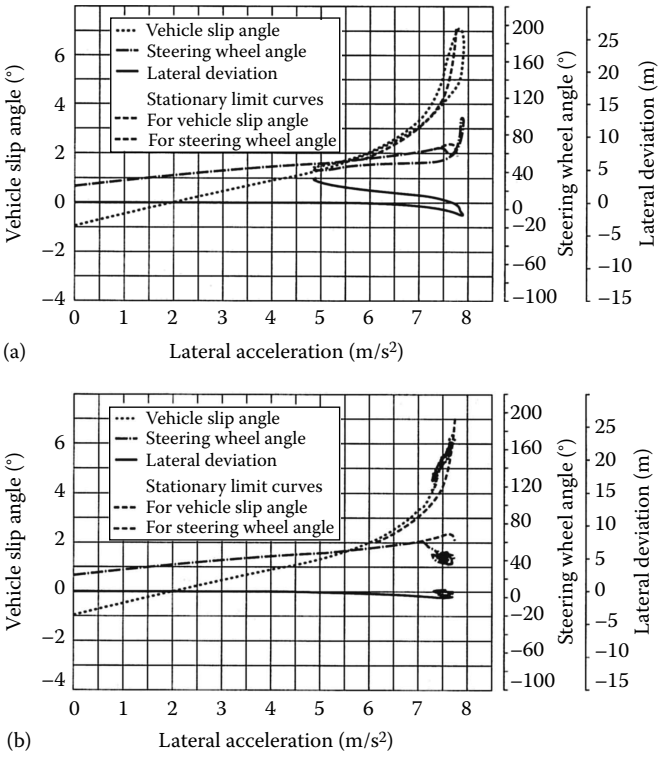


FIGURE 30.51 Comparison of the performance between a vehicle without (a) and one with (b) ESP for a steady-state cornering maneuver with constant radius of the turn and with slowly increasing velocity.

Up to a lateral acceleration of approximately 7 m/s^2, the behavior of the conventional vehicle and the ESP vehicle is identical and almost equal to the stationary behavior. Beyond this lateral acceleration value, the behavior of the conventional vehicle becomes different from that of the ESP vehicle and from the stationary behavior. The vehicle slip angle and the steering angle increase rapidly and progressively. At the lateral acceleration of 7.5 m/s^2, the conventional vehicle drifts out of the circle. Beyond the lateral acceleration value of 7 m/s^2, the ESP system intervenes such that the required steering angle values are reduced and such that the vehicle slip angle value stays at an almost constant value. Although the driver still gradually increases the accelerator pedal position value, the engine output torque is kept limited under the influence of the ESP and the vehicle speed is no longer increased since the physical limit of the vehicle motion is reached. The small lateral deviations of the vehicle from the track are reduced by the steering actions of the driver. As can be seen from the results, the driver's actions at the lateral acceleration value of 7.5 m/s^2 do not result in an unstable vehicle behavior. The resulting changes in the vehicle slip angle and the lateral deviations of the vehicle from the track both stay within a small range.

The simulation results were obtained using a modular vehicle dynamics simulation package called FASIM [45]. The simulation program has been validated and accurately reproduces the dynamic behavior of the test vehicle. The results presented earlier have been verified with the test vehicle.

Figure 30.52 shows the simulation results of a slalom maneuver in which the performance of a conventional vehicle without ESP is compared with that of the same vehicle with ESP. The steering wheel is rotated stepwise, in opposite directions, at regular time intervals and with increasing step values. The road surface is homogeneous with a medium value of the coefficient of friction ($\mu = 0.45$). During the maneuver, the accelerator pedal position is gradually increased in order to

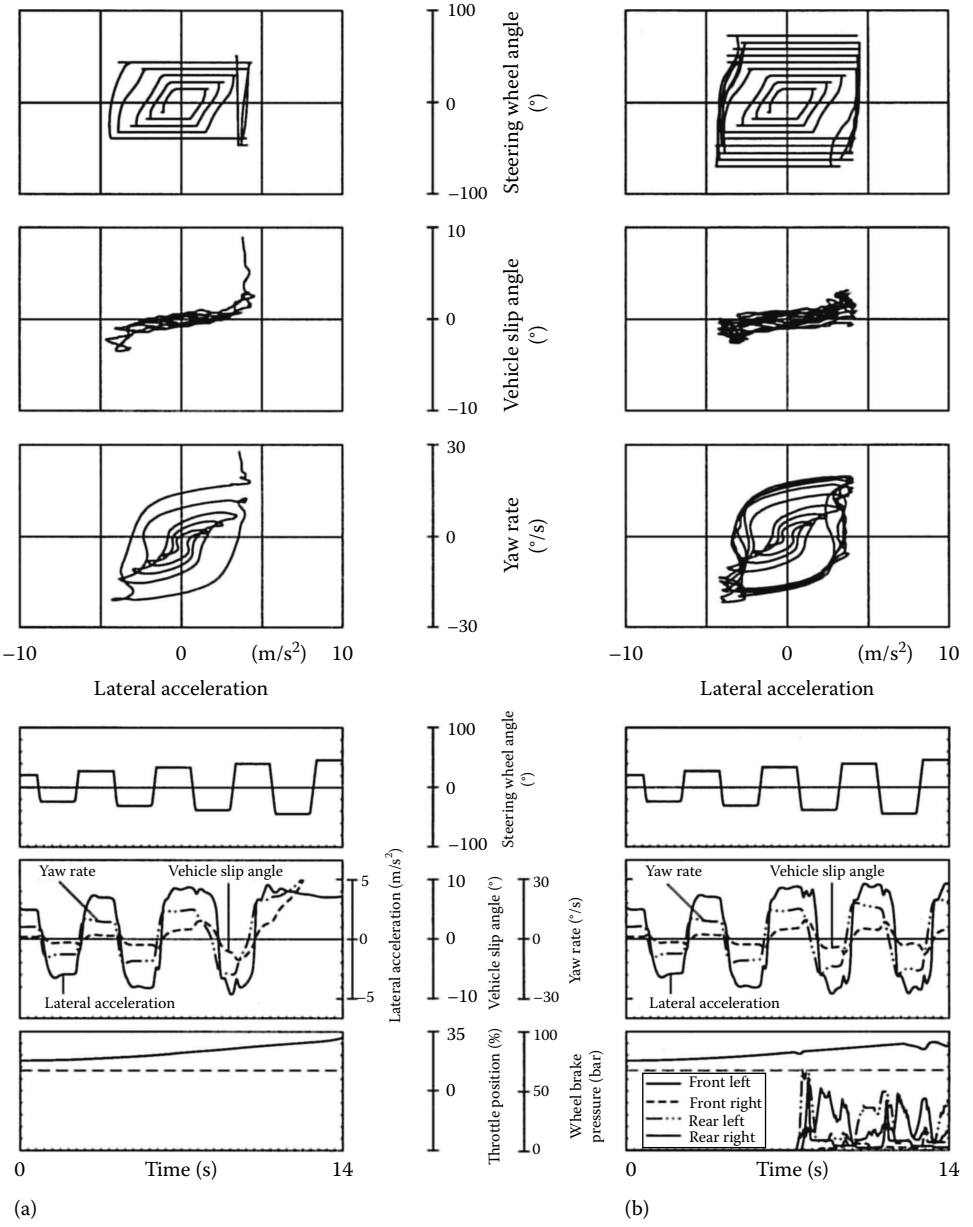


FIGURE 30.52 Comparison of the performance between a vehicle without (a) and one with (b) ESP for a slalom maneuver with increasing steering wheel angle amplitude and constant vehicle velocity.

keep an approximately constant vehicle speed. This will result in a steady increase in the drive slip at the driven rear wheels. On the left of Figure 30.52, the results of the conventional vehicle are depicted (Figure 30.52a), while on the right of Figure 30.52, the results of the vehicle with ESP are shown (Figure 30.52b). The upper portion of Figure 30.52 shows the results as phase plots depending on the lateral acceleration of the vehicle. The lower portion of Figure 30.52 shows the results as time histories.

The conventional vehicle becomes unstable after the fourth positive steering wheel step input. The steering wheel angle is then 40°. Both the yaw velocity and the vehicle slip angle increase rapidly while the lateral acceleration stays at an almost constant value. The results of the maneuver

with the vehicle with ESP show that the ESP system intervenes already after the third positive steering step input with a slight reduction in the value of the throttle valve angle and with active braking at the front and rear wheels in order to improve the steerability of the vehicle and in order to keep the vehicle stable. The vehicle motion also follows the sequentially increasing steering wheel step inputs stably, with values of the vehicle slip angle and yaw velocity that are limited by the physical properties of the environment.

30.2 SPECIAL FUNCTIONS OF ESP

30.2.1 Off-Road Vehicles

30.2.1.1 ABS Braking Off Road

The extensions of four-wheel drive vehicles to off-road vehicles refer to the special environment in which these vehicles may be operated. In particular, the road surface may be loose like a sandy surface. One extension of ESP with reference to this aspect is discussed here [46].

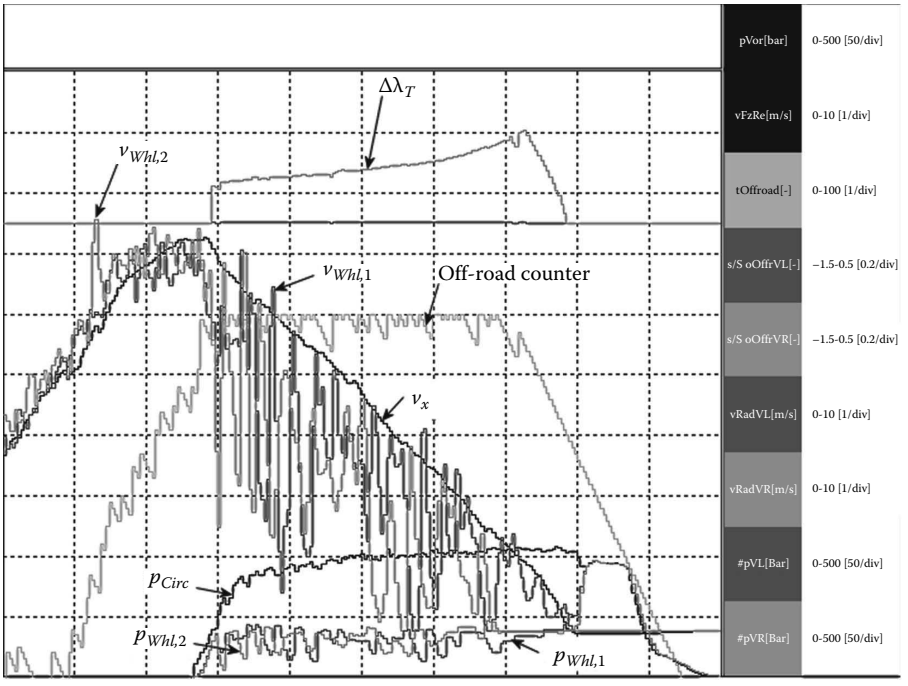
It is well known that on loose road surfaces like deep sand (and deep snow), the maximum brake and traction forces are achieved at large slip levels, that is, slip levels that are much larger than those on asphalt road surfaces. The brake coefficient of friction of a wheel in deep sand whose slip is λ_0 of dry asphalt (e.g., $\lambda_0 = 0.1$) may be about 0.3, while that of the locked wheel may be 0.7. For improved stopping distances, it would be of great advantage if during ABS the average slip could be increased. This is easily implemented by using the BSC of ESP. The extension of ESP for off-road ABS braking is therefore straightforward.

First, it must be detected that the vehicle is driving in off-road conditions. For this purpose, an off-road counter is introduced. If the off-road counter reaches a certain threshold value, then an off-road condition is expected. The off-road counter is increased if the wheel velocity oscillates with high frequencies. The larger the amplitude of the high frequency oscillations, the larger is the increment in the off-road counter. If the high frequency oscillations in the wheel velocity vanish, then the off-road counter is decremented.

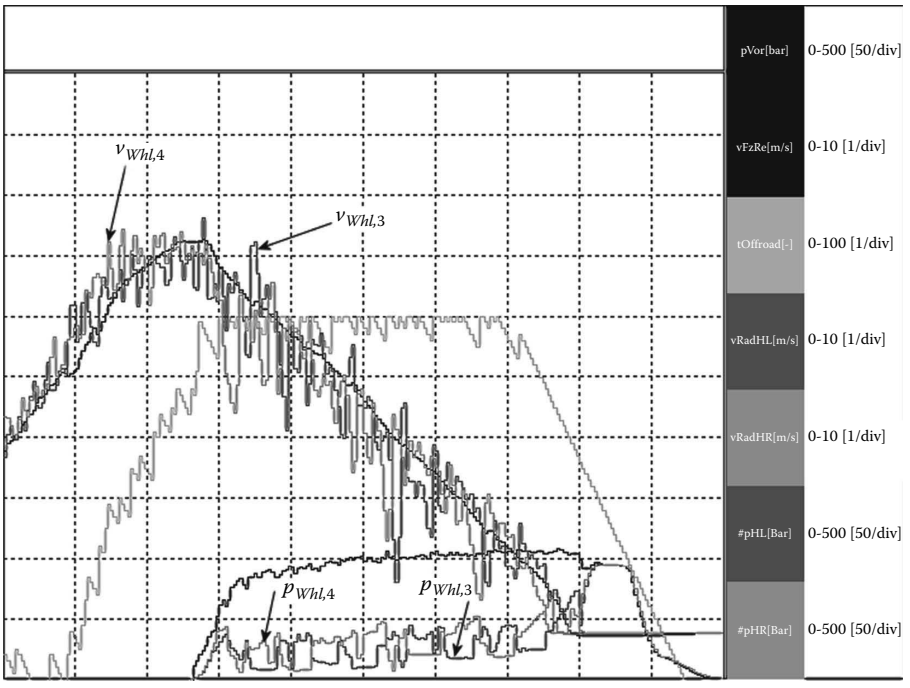
As soon as the off-road condition is expected, the target value of the ABS slip is increased at both front wheels. If it is not very clear that the off-road condition is valid, then the slip is increased at one front wheel only. There may be some ambiguity between braking on a packed snow surface and braking on deep sand, because for a small wheel slip the coefficients of friction are approximately equal. To differentiate between the two surfaces, the combination of slip and coefficient of friction of the road is investigated. If the wheel slip is large and the coefficient of friction is small, then the road surface is assumed to be packed snow or ice. If the wheel slip is large and the coefficient of friction is also large, then the road surface is assumed to be off road. If the distinction is not so clear, then the target slip is incremented at only one front wheel. The target slips at the rear wheels are left unmodified.

Figure 30.53 shows such a case in which it was not clear if the environment is off road or if the road surface is packed snow or scraped ice. At the beginning of the plot, the car accelerates and there are clear oscillations at the front and at the rear wheel velocities, particularly at the right front wheel. During the acceleration, the off-road counter is incremented and decremented as the amplitude of the oscillation increases and decreases, respectively, but in average the counter value increases. At the time the ABS braking maneuver starts (at 27 km/h), the counter has reached its threshold value and the target slip at the right front wheel is incremented initially by about $\Delta\lambda_T = 15\%$ and toward the end of the braking maneuver by about 30%. The increment $\Delta\lambda_T$ is added to the target slip $\lambda_{T,2}$ found by (30.22), so that the resulting target slip is set to about 25% at the beginning and to about 45% at the end of the braking maneuver.

The feature is available only at vehicle velocities below 40 km/h in order to keep the vehicle stable and steerable at larger velocities. Furthermore, the slip increment is reduced to zero immediately if the driver turns the steering wheel.



(a)



(b)

FIGURE 30.53 Time histories of some variables at the front wheels (a) and at the rear wheels (b) of a four-wheel drive vehicle during an ABS straight-line braking maneuver with off-road conditions.

30.2.1.2 Hill Descent Control

Many off-road vehicles have an additional countershaft transmission with which the total transmission ratio can be increased and which can be switched on or off depending on off-road conditions. The gear can be specially used for hill descent situations. The driver does not have to continuously brake by going downhill in off-road situations and he can pay full attention to the task of steering the vehicle. Other off-road vehicles may not have the additional countershaft transmission. For those vehicles, ESP has a feature called "Hill Descent Control" with which the vehicle can be braked automatically so that the driver can again pay full attention to the steering task [46].

The feature HDC can be activated by the driver by pushing the HDC button. The feature is activated only if the vehicle is moving downhill. Since the slope of the road is already obtained by the estimation of the vehicle velocity, using the Kalman filter (30.21), the implementation of the feature is also straightforward. The larger the offset, the larger is the slope of the road. A key factor for the controlled downhill braking is the information on the road slope. The feature keeps the vehicle velocity at a constant value of 8 km/h by automatically applying the brakes at all wheels. By active braking and modulation of the brake pressures, the vehicle velocity is kept constant depending on the road surface friction. It is possible that the road friction is locally so low that the wheel may start to lock during the hill descent braking control. In such cases, the BSC will limit the wheel slip to the nominal slip value on top of the hill descent control. This means that an ABS-braking maneuver may be in operation, although the driver does not apply the brakes. The coefficients of friction of off-road surfaces are typically in the range of 0.2–0.6. As reported earlier, on deep sand roads (also deep snow and deep gravel), the coefficient of friction may reach values (depending on the slip value) between 0.3 and 0.7.

30.2.2 Rollover Mitigation

The standard tuning of ESP is usually sufficient to reduce the danger of rollover for all cars [47]. However, for vehicles with a soft suspension and with a high center of mass, the special function "ROM" may be activated to even further reduce the rollover instability. This function contains a prediction of an incipient vehicle rollover. The prediction is based on the rate of change of both the steering angle and the lateral acceleration of the vehicle. If an incipient rollover is predicted, then the front wheel on the outside of the turn is actively braked to reduce the lateral acceleration of the vehicle. Braking will also reduce the vehicle velocity and with that the centripetal acceleration and the threat of rollover.

Other systems may use a second yaw rate sensor to measure the rotational velocity of the vehicle about its longitudinal axis (x-axis) [48]. Both the rotational velocity and the rotation angle, which is obtained by the integration of the rotational velocity, are used to predict the danger of rollover. If a rollover danger exists, both wheels on the outside of the turn are automatically braked.

30.2.3 Trailer Oscillation Mitigation

A well-known situation for a vehicle towing a trailer is that the trailer can spontaneously start to oscillate about the hitch of the towing vehicle and about the vertical axis (z-axis) of the trailer. The situation arises if the velocity of the vehicle–trailer combination lies above the so-called critical velocity. If the velocity of the combination is below the critical velocity, then the oscillations about the z-axis are damped. Using a simple bicycle model, it can easily be shown that the damping ratio has the vehicle velocity in its denominator. This means that the damping of the oscillation depends on the vehicle velocity. The damping ratio decreases, the closer the velocity approaches the critical velocity. If the velocity is above the critical velocity, the damping ratio becomes positive, the trailer oscillation becomes unstable, and the amplitude of the oscillation increases. Drivers of such combinations are recommended to firmly apply the brake pedal and slow down the vehicle. If the vehicle is slowed down to below the critical velocity, the oscillation will vanish by itself.

ESP contains a feature to recognize the oscillation in the trailer [49]. If the trailer oscillation continues for some periods and does not vanish by itself, then the brakes are applied automatically. The recognition of the trailer oscillation is possible since the oscillations of the trailer can be seen in the yaw velocity of the towing vehicle. It takes about 3–6 periods of the trailer oscillation to recognize the oscillation of the trailer by analysis of the vehicle yaw velocity signal. The brakes are applied by specifying the vehicle deceleration and adjusting the brake pressures by the hydraulic model. Two deceleration levels are used. If the amplitude of the oscillation is large, then the large level (0.5 g) is used. Otherwise the small level is used (0.3 g). If the amplitude of the yaw velocity oscillation of the towing vehicle becomes very large, ESP may intervene ahead of the deceleration command to reduce the yaw velocity. Additionally, ahead of the deceleration braking, the braking of the towing vehicle may be asymmetric to actively damp the oscillation.

30.2.4 Electronic Brake Force Distribution

Brake regulations require that the brakes are dimensioned such that the front wheels lock before the rear wheels lock. If the vehicle deceleration is larger than 0.85 g, then the requirement does not apply. Since the dimensioning of the wheel brakes is fixed, there is a linear relationship between the brake forces on the front axle and on the rear axle. The result is that usually the rear wheels do not contribute much to the deceleration of the vehicle.

The situation is shown in Figure 30.54. The brake force distribution is called "ideal" if the coefficients of friction of the wheels at both axles are equal. The fixed brake force distribution cuts the ideal brake force distribution at the vehicle deceleration of 0.85 g. For this fixed brake force distribution, the rear brake force is much smaller than the brake force of the ideal brake force distribution, in particular for small brake forces, that is, small vehicle decelerations. That means, for small vehicle decelerations (i.e., for by far the most brake applies), the brake forces at the rear wheels according to the fixed brake force distribution contribute much less to the deceleration than the front wheel brakes do. In order to improve the situation for small decelerations, the line of the fixed brake force distribution can be made steeper. But then the line cuts the ideal brake force distribution at a much lower vehicle deceleration (e.g., at 0.5 g), which is not allowed by the brake regulations. In order to still comply with the brake regulations, the ABS is used to reduce the brake force at the rear axle if the deceleration is larger than at this intersection. This function is called EBD. Although the brake pressure at the rear axle is not measured, this is still possible by control of the velocities of the rear wheels.

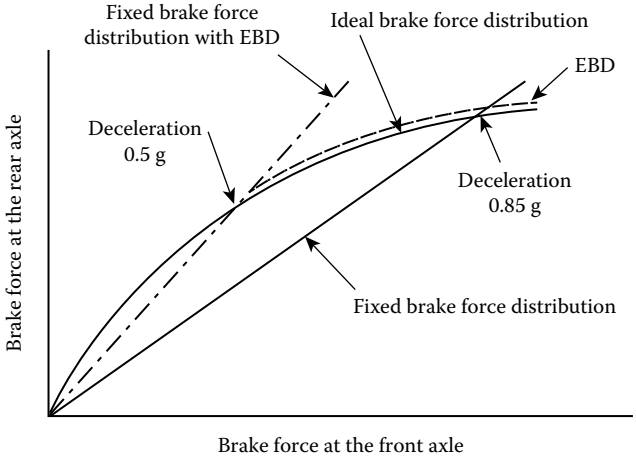


FIGURE 30.54 Principle of a fixed brake force distribution in combination with the EBD.

Along the ideal brake force distribution curve, the coefficients of friction of the wheels at both axles are the same. Since the road surface is homogeneous, the slips of the wheels at both axles are then also equal. But this means that along the ideal brake force distribution curve, the wheel velocities at the rear axle equal those at the front axle. If the wheel velocities at the rear axle are smaller than those at the front axle, then the slips at the rear axle are larger than those at the front axle and then the coefficient of friction at the rear axle is larger than that at the front axle and then the rear wheels lock first. It is therefore sufficient to monitor and control the difference in wheel velocities between the front wheels and the rear wheels and to keep the velocities of the rear wheels larger than the velocities of the front wheels. For best utilization of the rear wheel brakes, the velocities of the rear wheels should be close to the velocities of the front wheels.

In fact, the velocities of the rear wheels are kept slightly below the velocities of the front wheels. As soon as the velocities at the rear wheels become smaller than those at the front axle and the difference is larger than the first threshold (which depends on the vehicle velocity), the brake pressure at the rear axle is no more increased. If the velocities of the rear wheels are smaller than the velocities at the front axle and the difference becomes larger than a second threshold, then the brake pressure at the rear axle is so much reduced that the difference becomes smaller than the second threshold. If the velocities of the rear wheels are smaller than the velocities of the front wheels but the difference is smaller than the first threshold, then the pressure in the rear wheel brakes is increased in a stepwise manner.

If a failure in the system prevents the EBD function and a vehicle deceleration of 5.8 m/s cannot be achieved without the locking of the rear wheels, then the EBD warning light in the dashboard should be lit.

30.2.5 Brake Assistant

The brake assistant (BA) belongs to the so-called driver assistance functions. Since its introduction, the number of assistance functions has increased steadily (see earlier text). The idea of the assistance functions is to assist the driver operating his vehicle and to improve the traffic safety in general. The goal of the BA is to help him in panic-braking situations and to reduce the stopping distance.

Investigations by Mercedes using a full-size driving simulator have shown that in panic situations many average drivers apply the brake with some period of hesitation at the beginning (Figure 30.55) [50]. This hesitation has a large influence on the stopping distance since it occurs at the highest vehicle velocity. If the panic situation is recognized by the BA, the pressure in the brake BCs is automatically increased during the hesitation period to a higher level from that commanded by the driver. Together with the assistant, the average driver can then perform an emergency-braking maneuver as well as an expert driver can do.

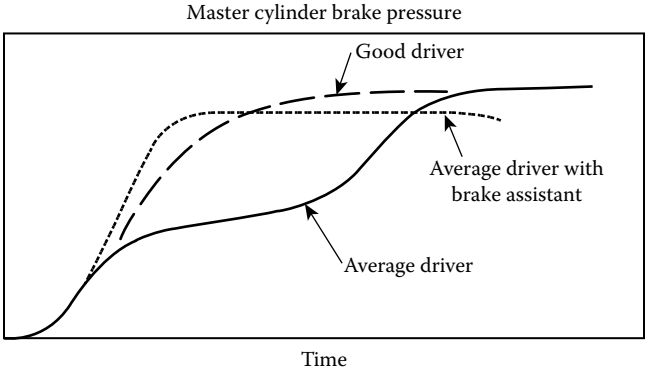


FIGURE 30.55 Principle of the support of an average driver using a pneumatic brake assistant.

The most important functional requirements of the BA are the following.

- Support of the driver in emergency brake situations and reduction in the stopping distance to values that can be achieved by well-trained drivers.
- Elimination of the support if the driver reduces the force on the brake pedal significantly.
- To leave the conventional brake booster function unchanged. Brake pedal feel and comfort must comply with the usual standard during normal braking.
- Activation of the BA only in real emergency situations such that the driver does not adjust his habitual behavior to the BA.
- If the BA is switched off, the conventional brake must not be impaired.

The key task of the BA is to recognize the emergency situation on the basis of the driver behavior. The recognition depends on the type of the BA. There are three types of BAs:

- The pneumatic mechanic BA called "emergency valve assistant" (EVA)
- The pneumatic electronic BA (PBA)
- The hydraulic electronic BA (HBA)

The EVA is realized by an extended design of the vacuum brake booster. Depending on the velocity with which the brake pedal is applied, the gain of the booster is switched to a larger value.

The PBA is realized by using a smart vacuum booster. Using a magnetic valve, the smart booster can allow air to enter the booster up to the maximum support for the driver. For the analysis of an emergency situation, the movement of the booster membrane is measured. The BA support is switched off if the membrane has moved back again and has pushed a release switch.

The HBA will be discussed now in some more detail [51]. The concept of the hydraulic brake assistant is shown in Figure 30.56.

Using the pressure sensor of ESP, the BA function is activated in phase 1 or deactivated in phase 2 according to the logic of Table 30.2. The thresholds can be chosen to adapt the function to the vehicle and its brake system. Furthermore, the thresholds are modified depending on the vehicle velocity, the brake MC pressure, the state variables of the BSC, and the analysis of the time history of the brake maneuver.

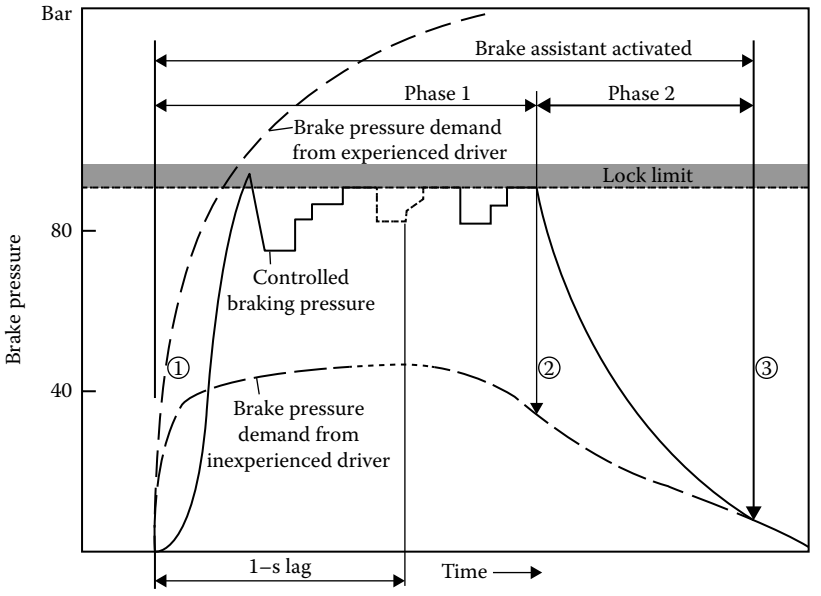


FIGURE 30.56 Principle of the hydraulic brake assistant.

TABLE 30.2
Concept of the Hydraulic Brake Assistant

Situation	Detection Logic
Phase 1 (Figure 30.56) Emergency situation Panic braking	Brake pedal applied *and* MC-pressure gradient above threshold-1 *and* MC-pressure above threshold-2 *and* vehicle velocity above threshold-3
Phase 2 (Figure 30.56) Reduced brake demand	Pedal force (derived from MC-pressure) below threshold-4
Renewed activation	MC-pressure gradient above threshold-1
Standard braking	Brake pedal not applied *or* MC-pressure below switch-off threshold-1 *or* vehicle velocity below switch-off threshold-2 *or* pedal force large enough

As soon as the emergency situation is detected, the BA support is activated (number 1 in phase 1). The BA function increases the brake pressure at all wheels until ABS control starts using the recirculation pump of the HU of ESP. The control situation is comparable with the ESP brake control interventions during partial braking. If the force on the brake pedal is reduced such that the brake pressure in the MC drops below threshold −4 (number 2), then the brake pressure support of the BA is reduced (phase 2). The control strategy is changed now. The goal is then to make the brake pressure at the wheels follow the pressure in the MC and to find a comfortable transition to a standard brake situation. The BA is switched off as soon as the increased brake pressure reaches the value of the MC or if the MC pressure has reached the switch-off threshold. In contrast with the PBA, all wheels are in ABS control while the support of the PBA may not be sufficient enough to reach that situation.

30.3 ESP SAFETY CONCEPT

ESP is a complex mechatronic system because of the large number of sensors and actuators, and the consideration of the vehicle and driver behavior (Figure 30.57) [32]). The ESP system requires information from several sources (four-wheel velocities, yaw velocity, lateral acceleration, steering wheel angle, brake pressure, brake light switch, hand brake switch, and a few signals from the engine and

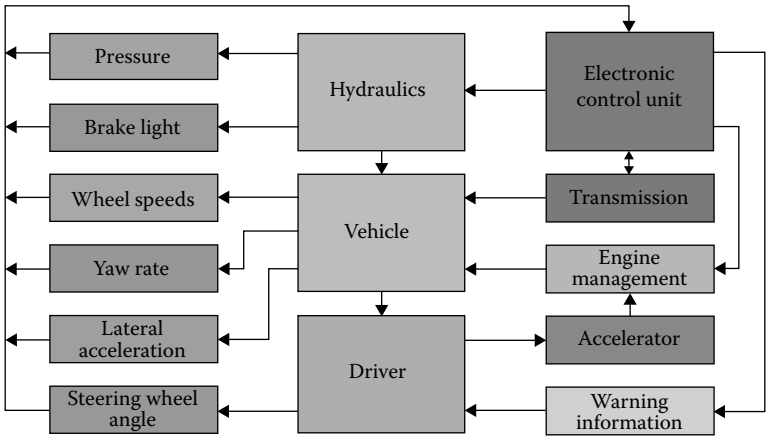


FIGURE 30.57 System components included in the monitoring concept of ESP.

transmission controller). Based on this information, the ESP controller calculates the signals for up to twenty-one actuators (12 valves, pump motor, engine and transmission control, up to three driver information lamps, the optional buzzer, and the lamp in the ESP ON/OFF switch), which modify the wheel forces or inform the driver. In the case of a failure, the ESP system must not use these components improperly. The main purpose of the ESP safety concept is to avoid such malfunctions.

30.3.1 Requirements of the Safety System

- The behavior of the vehicle with ESP must not be worse in any situation as compared with a vehicle without the system (i.e., component failures must not lead to dangerous driving situations).
- System failures must be detected sufficiently fast so that the safety of the passengers and other traffic participants is not impaired.
- Only one front wheel may be controlled to very high slip values at the same time, so that in an emergency situation the other front wheel can be used for handling maneuvers.
- After a system failure has occurred, the best possible switch-off strategy must be chosen as defined by the vehicle manufacturer.
- If a system failure is detected during an ESP intervention, the control must continue with a limp home functionality up to the end of the ESP-relevant situation.
- ESP must provide fallback functions. It is required that the ABS function is kept alive as long as possible, preferably also if only four WSSs are available.
- With some faults of short duration (like electromagnetic interference), the system can be kept activated if the control performance is still acceptable. Otherwise, if the control performance is reduced repeatedly or reduced for a longer period of time, then the system must be switched off.
- The control may be activated only after the safety system has released the control system.

The current status of the basic principle of the ESP safety concept is (Figure 30.58):

- Failure avoidance by analysis of potential problems before the effects influence the controller
- Self-monitoring and self-test functions for key sensors
- A complex and sophisticated monitoring logic which can detect components which are slightly beyond specification and which can decrease the controller sensitivity in the case of suspicion of a failure
- Gradual or immediate shutdown of the system depending on the type of failure
- Limp home functionality depending on the type of failure

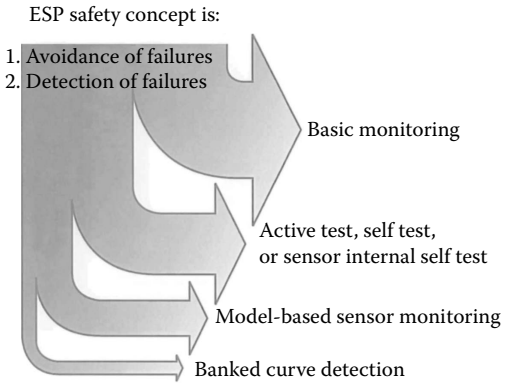


FIGURE 30.58 ESP safety concept.

30.3.2 Avoidance of Failures

Experience with the first ESP implementations has shown that an intensive and early discussion between the supplier and the customer about the safety concept of the ESP system is essential. Only a comprehensive approach to all possible influences during the whole life of the vehicle and a thorough discussion between the customer and the supplier can sufficiently reduce the risk of malfunction of the system or the components. Failures and their effects caused by abuse, modifications in the specification of shared components, and/or their use, mistake or damage during a service, or as a result of an accident must be considered. For this reason, problem-solving analysis using the FMEA principle (failure mode and effect analysis [FMEA]) has been implemented early in the development programs. In these extensive discussions, misunderstandings or vagueness of definitions can be prevented; on the other hand, the customer is very well instructed about ESP and his requirements are integrated into the safety concept. In case the customer contributes important components, for example, the brake MC, the WSSs, or the engine and transmission controller, not only the definition of a functional interface is necessary but also a comprehensive consideration of the safety techniques and equipment concerning these components is essential. Since in the end, the safety and the availability of ESP are determined by the weakest link of the whole mechatronic system; this effort is always justified and worthwhile.

Most suppliers use FMEA to analyze the possible problems. The first FMEA meeting starts as early as possible and continues during the course of the project. As a result, changes and commonly defined measures can be incorporated in the design in a timely manner. The FMEA produces a guide to realize improvements and to avoid failures. These considerations and discussions of the system by means of the FMEA go deep into the details of the system and are a cornerstone of the safety concept.

An important aspect of the FMEA is the judgment of the severity of the effect of failures. Here are some examples of the severest effects (rated 10 and 9):

1. Unintentional braking of one or more wheels outside ESP operation, which leads to a clear lateral motion of the vehicle.
2. Unintentional acceleration of the vehicle (>1.5 m/s^2).
3. Unintentional deceleration of the vehicle (>2.0 m/s^2).
4. Impossibility to brake (e.g., all EVs are closed).
5. Fire or smoke.
6. The warning light is not lit despite a detected failure.
7. The brake lights are not lit although the driver applies the brake.
8. The minimal deceleration required by law cannot be achieved.
9. Unintentional braking of one front wheel during TCS operation.
10. ESP is switched off or its performance reduced during an intervention or shortly before an intervention is necessary, which leads to a drifting or swerving vehicle.
11. Significant reduction in the brake efficiency:
    a. Stopping distance increased by more than 40%.
    b. No brake pressure at the front axle.
    c. No brake pressure at the rear axle.
    d. No brake pressure in one brake circuit.
    e. No brake pressure at one front wheel.

Here are some examples of severe effects (rated 8 and 7):

1. Unintentional braking of one or more wheels outside ESP operation, which is noticed by the driver but which does not lead to a clear lateral motion of the vehicle.
2. Unintentional acceleration of the vehicle (>1.2 m/s^2 but <1.5 m/s^2).
3. Unintentional deceleration of the vehicle (>1.5 m/s^2 but <2.0 m/s^2).
4. Rear wheels are locked during braking.

5. Unintentional braking of one driven wheel during TCS operation.
6. ESP is switched off or its performance is reduced during an intervention or shortly before an intervention becomes necessary, which leads to an insecure feeling in the driver, although the vehicle does not clearly drift or swerve.
7. Reduction in the ESP performance, which causes fear in the driver:
   a. Stopping distance increased, however, less than 40%.
   b. No brake pressure at one rear wheel.

30.3.3 Self-Test, Self-Monitoring, and Tests of Components

Since sensor redundancy is expensive, self-test, self-monitoring, active tests, and analytical redundancy are used for sensor monitoring. Besides the yaw rate sensor, the steering wheel angle sensor is very important to the ESP system. Since the measurement of angles is state of the art and can be done in a very robust and reliable way, the steering wheel angle sensor has been defined as the "master sensor" of the system. This means that the sensor must be able to detect internal failures by itself and to report any failure to the ECU. The time granted to recognize failures is fitted to the requirements of the ESP system and is less than 100 ms. If possible, the system uses a robust digital interface (CAN bus) to transfer data from the sensor to the control unit. Only if the chosen steering wheel angle sensor meets these demands, will the safety algorithms, especially the model-based sensor monitoring, detect the occurring failures in time and with the necessary accuracy and identify the defective component.

Self-test functions are also integrated in the yaw rate sensor, the lateral acceleration sensor, and the brake pressure sensor. The yaw rate sensor is provided with a built-in self-test (BITE), which is triggered externally by the ECU every 40 ms. Because of the test, the yaw velocity can be measured only every 40 ms. Figure 30.59 shows the signal of the yaw rate sensor. During the test mode, the externally stimulated self-test function (from the ECU) causes a superimposed bipolar test signal of ±25°/s at the output to the yaw velocity signal, after the transients have died out. The tolerance of this test signal is ± 9°/s. The analog-to-digital (A/D) converter samples the yaw velocity signal every 1 ms, and then with a delay of 20 ms after the stimulation the response to the test input is calculated by using the last three measured yaw velocity values. This filtering is done in order to take into account the changes in the yaw velocity signal itself as precisely as possible, during high dynamic driving conditions also. Because of the robust evaluation during these three cycles, a very fast reaction of the system is possible in case of a detected failure.

By analogy to the yaw rate sensor's BITE, the lateral acceleration sensor contains a similar test function, which however is fully integrated in the sensor. In case of a failure, the sensor signal is switched into the out-of-range band, which can quickly and safely be classified as a defective state of the sensor.

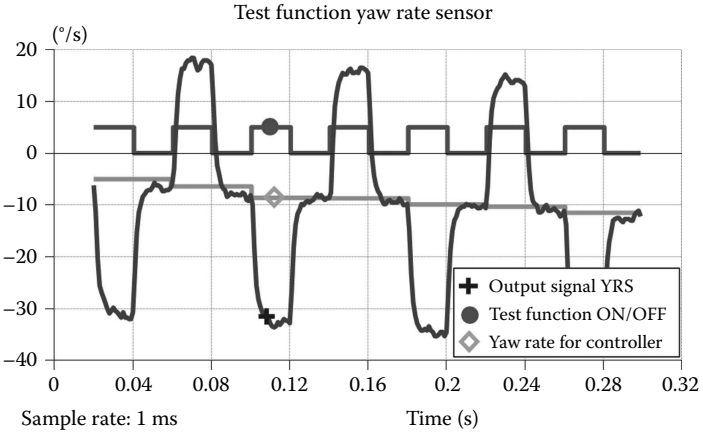


FIGURE 30.59 Yaw velocity signal including the self-test function.

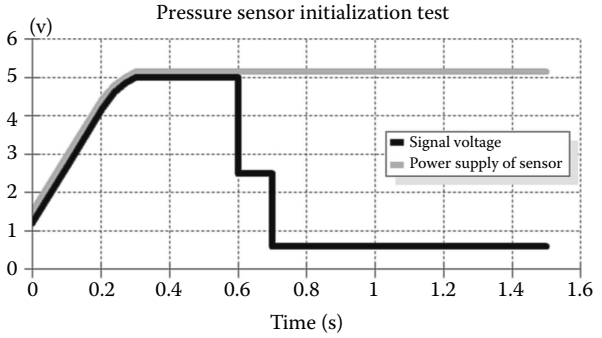


FIGURE 30.60 Power on test of the pressure sensor.

The brake pressure sensor is provided with an internal monitoring of its sensitivity and offset, which in case of a failure pulls the output signal into the out-of-range band. In addition, the sensor is provided with an initial self-test in order to detect failures due to resistance of connectors and to check the functionality of monitoring. Figure 30.60 shows the output signal of the brake pressure sensor during the build up of the supply voltage. The output runs parallel to the supply voltage up to its maximum value of 5 V. About 300 ms later, the sensor must show a predefined value of 2.5 V for a period of approximately 100 ms. Only then does it show the measurement signal.

Besides these self-test functions of the ESP sensors, active tests can be realized by means of the complex hydraulic system. Its components can be activated and the results of this activation can be examined by the sensors. For example, the precharge pump can be tested if it is switched on while the EVs of the HU are closed and the increase in the pressure in the push rod brake circuit due to the activated pump is examined (Figure 30.70). At the same time, using the same procedure, a crude test of the pressure sensor can be obtained. Even the hydraulic system can be tested and it is possible to distinguish between the two brake circuits by measuring the spin downtime of the recirculation pump for each of the brake circuits. By this procedure, called "active hydraulic test," it is possible to detect mechanical defects in the hydraulic system and to prevent an active pressure build up in case of a failure. It should be noticed, however, that this test significantly increases the stress on the HU during its lifetime.

30.3.4 Failure Detection Logic

The failure detection logic is divided into the following parts (Figure 30.58):

- Basic monitoring
- Evaluation of the active tests
- Self-test functions
- Model-based sensor monitoring
- Measures which reduce the effect of unrecognized failures

The first part, the basic monitoring, includes most of the well-known monitoring logic of ABS and TCS. It monitors the electrical functions of the valves and the pump motor, the correct function of the WSSs, the power supply for the ECU and the additional sensors, the electronic components inside the ECU, the wires and the connectors to the ESP sensors, and the CAN bus. The logic is extended to wheel speed monitoring because the ESP controller uses very low thresholds for the slip controller. Improved failure detection logic was necessary in order to prevent malfunction of the system.

The second part of the safety logic is the evaluation of the self-test functions of the yaw rate sensor and the pressure sensor as described earlier and the realization of the active tests. The other parts will be discussed in the following sections.

30.3.4.1 Model-Based Sensor Monitoring

The model-based sensor monitoring is used for two purposes. First, it is used to improve the fault tolerance of ESP, for example, errors in the sensor signals do not reduce the system performance. Second, it is used to detect failures of the sensors. The concept of the model-based sensor monitoring is discussed here on the basis of [52]. Further discussions can be found in [53,54].

30.3.4.1.1 Fault Tolerance of ESP

In order to satisfy the requirement that the system must be fault tolerant, the sensors are continuously calibrated if the ignition is on. The steering wheel angle and the lateral acceleration signals are calibrated based on the assumption that on average every car moves straight ahead. Thus, the offset of the steering wheel angle signal and the lateral acceleration signal are obtained by very low pass filtering of each signal. Special care must be taken if the driver steers continuously in one direction, for example, as is the case during high-speed test maneuvers on special tracks at proving grounds. In this case, the offset filtering must be interrupted.

The offset of the yaw rate sensor is computed based on the assumption that if the vehicle stands still then the yaw velocity is zero. Thus, the measured yaw velocity signal at vehicle standstill is the offset of the signal. Special care must be taken at "zero" vehicle velocity. Since, using passive WSSs, the wheel velocity cannot be measured; if the velocity of the wheel is smaller than 2.75 km/h, the car may creep forward but the measured wheel velocities are zero. If the steering wheel is turned all the way up to its limit (usually ± 720°), then the yaw velocity may be high (e.g., 10°/s) and this value must not be interpreted as the signal offset. Furthermore, there are parking buildings which use an elevator for the vehicles to transport them to the different floors (e.g., in Tokyo, Japan). These elevators can rotate, for example, by 90°, to let the vehicle out and this rotation is measured as the yaw velocity of the vehicle, although the vehicle is really standing still.

In addition to the offset compensation of the yaw velocity during vehicle standstill, there is the possibility to determine the offset during driving. For this offset calibration, a reference yaw velocity must be determined whose value is very reliable and then the measured yaw velocity can be compared with the reference yaw velocity to determine the yaw velocity offset. The reference yaw velocity is obtained by the "model-based sensor signal monitoring" (also called "analytic redundancy"). The basic principle of this part of the safety concept is shown in Figure 30.61. This monitoring method works very well and even sensitivity changes of the yaw rate sensor can be determined.

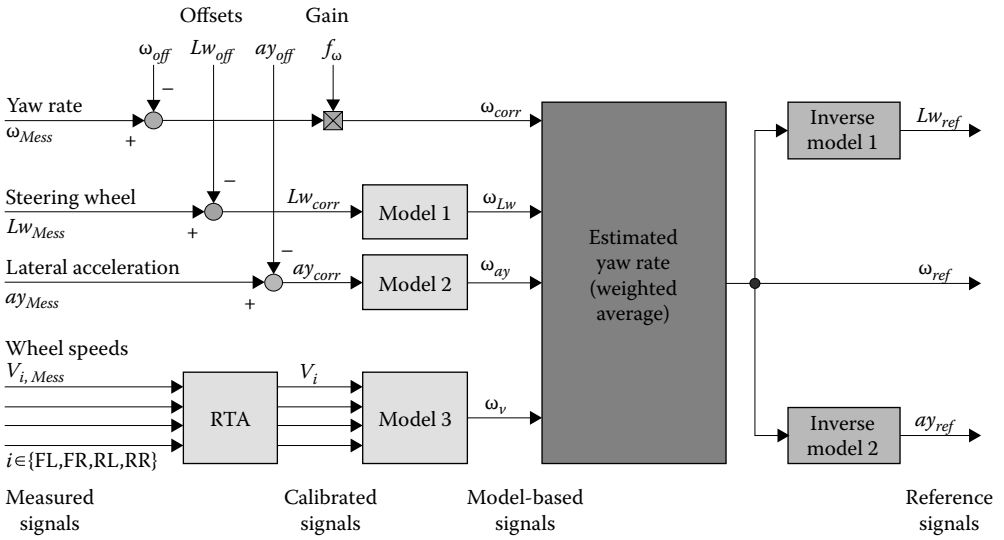


FIGURE 30.61 Model-based monitoring of the sensor signals.

The measured signals from the sensors for yaw velocity, lateral acceleration, steering wheel angle, and wheel velocity are offset corrected. An additional sensitivity correction of the yaw velocity signal is implemented. With the corrected steering wheel angle, lateral acceleration, and wheel velocity signals, yaw velocity estimates can be obtained by using several models. Model 1 is the linear steady-state bicycle model (Table 30.1). Model 2 is Equation 30.5. Model 3 can be found from the difference in the velocities of the wheels at one axle divided by the track width. It is important to have the wheel velocities calibrated first (the calibration is called RTA) before model 3 can be applied. Because the estimates depend on the validity of the models used, the yaw velocity can only be estimated if the vehicle motion is well within the physical limits (i.e., if the slip angles at the wheels are well within the linear region of the slip angle curves).

Together with the measured yaw velocity, four analytically redundant yaw velocity values are derived from which a reference yaw velocity value ω_{ref} is determined by computation of a weighted average of the four values. The reference yaw velocity obtained is insensitive to a sensor fault and corresponds with large probability to the actual vehicle yaw velocity. With the help of inverted models, it is possible to determine in addition to the reference yaw velocity, a reference steering wheel angle and a reference lateral acceleration. A residual is then built by comparing the reference signals to the measured signals. If the integrated residual grows too large, the ESP function is turned off. The model-based fault detection is so accurate that it is possible to detect absolute faults in the yaw velocity signal down to 3°/s.

The total procedure of determination of the offset and of the sensitivity change of the yaw velocity signal and the detection of sensor faults is shown in Figure 30.62.

The offset computation during vehicle standstill has been discussed earlier. Since it is not always possible to obtain the offset during vehicle standstill, a fast offset computation is performed in such situations during drive off. This will result in the first crude information on the offset. Input to the fast offset computation is the difference between the sensor signal ω_{Mess} and the reference signal ω_{ref}. During normal driving the normal offset computation runs. It is the same computation as the fast offset computation, however, with a large time constant. Also, the sensor sensitivity is calibrated during normal driving. For the sensitivity calibration, large yaw velocity values are required as is the case for driving in urban areas. The vehicle dynamics control starts with the ignition on, with enlarged dead zones of the actuating signals of the yaw velocity and body slip angle; and, only if a precise estimate of the offset is obtained, the dead zones are reduced to the normal values. Offset and sensitivity are stored in EEPROM.

As long as the models shown in Figure 30.61 are valid, precise estimates of the offset and sensitivity of the yaw velocity signal can be obtained. Figure 30.63 shows how often the models are valid during driving on different roads and conditions. It shows that the models are most of the time valid while driving on normal roads, but that the situation is different during driving in the mountains with sharp and steep turns.

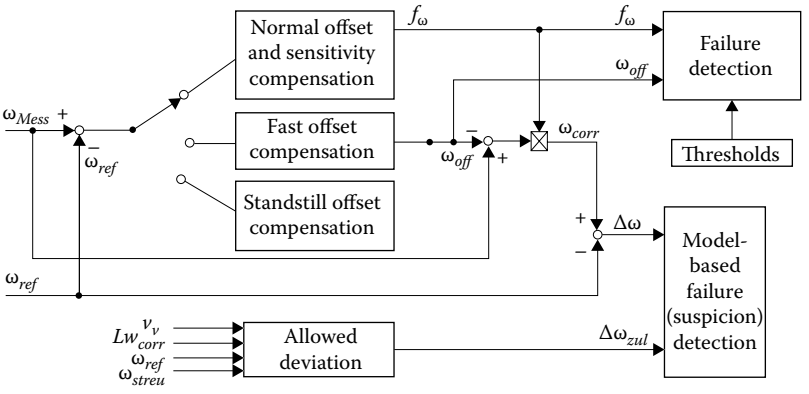


FIGURE 30.62 Calibration and fault detection in the yaw velocity signal.

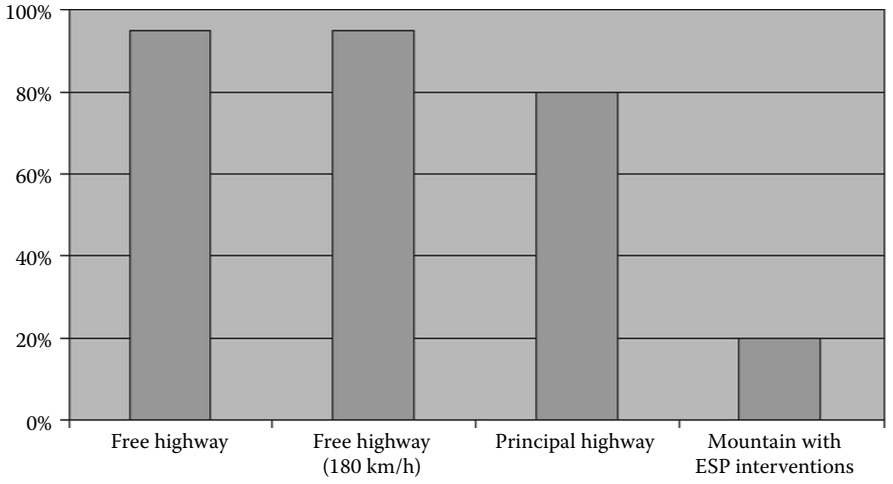


FIGURE 30.63 Validity of the models in percent of the traveling time on various roads.

30.3.4.1.2 Detection of Failures

Another purpose of the model-based fault detection is to deactivate the vehicle dynamics controller if a sensor failure occurs. For example, if a noticeable deviation of the measured yaw velocity or lateral acceleration from their reference values occurs, this will result in a suspicion of a sensor failure. The model-based fault detection will decide within 250 ms if the expected control intervention is allowed or if a faulty sensor signal is present and the controller is to be deactivated. In a very short time, the model-based sensor fault detection locates the failure, deactivates the ESP system, and turns the driver information lamp on.

The model-based failure detection of the yaw rate sensor will be discussed using Figure 30.62.

First, the computed offset and sensitivity faults are compared with the allowed thresholds and a failure is detected if their values are larger than those thresholds.

Second, the calibrated signal ω_{corr} is compared with the reference signal ω_{ref} and the difference $\Delta\omega$ is compared with a variable threshold $\Delta\omega_{zul}$. Now, if the difference is larger than the threshold, then the path chosen in the failure detection system depends on the driving situation.

- If the models are valid, then a failure is detected.
- If the models are not valid but the yaw velocity signal is "observable," then a failure might be present and since it is not sure if a failure has definitely happened only a "failure suspicion condition" flag is set. The calibrated yaw velocity sensor signal ω_{corr} is said to be observable if the computed yaw velocity values, ω_{Lw} from the steering angle, ω_{vR} from the wheel velocities, and ω_{ay} from the lateral acceleration, are close together in a narrow yaw velocity band for several sequential sampling points in time. So the models are still used for failure detection but not for the calibration of the sensor signal. If the failure suspicion condition flag is set, then the dead zones for the vehicle dynamics controller are enlarged in order to prevent unintentional yaw moment interventions. Depending on the magnitude of the difference $(\Delta\omega - \Delta\omega_{zul})$ and the time duration, the failure suspicion condition flag is set when a failure is detected.

The variable threshold $\Delta\omega_{zul}$ depends mainly on the specification value of the yaw rate sensor. Since the models are abstractions of the reality, the model-based yaw velocities also have some error. Therefore, the threshold must be larger than that defined in the specification. Furthermore, the model-based yaw velocities are derived from sensor signals that also have some error even after calibration. This is considered with the computation of the allowed error $\Delta\omega_{zul}$ by the value of ω_{streu}.

Based on the concept of using reference signals, it is also possible to improve the diagnostic function of ESP. If the basic monitoring or the evaluation of the self-test functions detects a malfunction, normally the influenced components are known. In case of model-based sensor monitoring, this is possible with the concept described. For instance, a difference between the yaw velocity signal and that derived from the steering wheel angle signal may show that there is a failure in the relationship of the signals and the system can react. But it is not possible to determine which of the two components is defective and to design an optimized shut-off strategy. An additional calculation would be necessary. By estimation of a reference signal, the monitoring can find differences in the signal path of the failed component and it is possible to store correct information into the EEPROM of the ESP ECU.

30.3.4.2 Measures for Unrecognized Failures

Failures that are not detected by the outlined detection functions could cause an unintended control intervention. But even for this unlikely case, consequence-restricting measures are taken. The most effective preventive measure appears to be the logic for preventing controller interventions during stable driving around banked turns. For instance, if the lateral accelerometer is stuck and continuously shows a zero acceleration value, then the combination of the yaw velocity signal and the lateral acceleration signal is similar to driving in a banked turn and similar to swerving on ice. The controller recognizes such situations by the described logic and prevents control interventions by compensation of the sensor signals (Figure 30.36). Thus, an unintended intervention is avoided in case the failure in the lateral accelerometer signal has not been detected yet.

Other plausibility checks reduce the length in time of the intervention. For instance, if the ABS control lasts implausibly long, then the control is stopped and the system is switched off. Also, if the brake light switch is set but the pressure sensor signal is zero for some length in time, the system is switched off.

30.3.5 System Behavior after Detection of Failure

The response of the ESP system after detection of a failure is based on the strategy which is regulated for the ABS and TCS systems. In the case of ECU and electrical failures of the hydraulic components, the whole system is shut off immediately and all warning lamps are switched on. In the case of WSS failures and pressure sensor failures, which occur while the control is active, it is possible to maintain some control with failure strategies until the end of the intervention. When failures of these components happen outside of active control, the ESP system and all its functions have to be shut down. In the case of failure of the yaw rate sensor, lateral acceleration sensor, or steering wheel angle sensor, control is switched over to a fallback ABS as described in the following text. The switchover to the fallback ABS function is indicated to the driver by switching on the ESP warning lamp.

After completely switching off the ESP system or the fallback ABS function, the EBD function or its limp home function is still available under certain circumstances. If even this function can no longer be maintained, the driver is informed by a red warning lamp (see the description of the EBD function).

30.3.6 Fallback Functions

The primary goal of very low failure rates of all ESP components in the field and therefore a very high availability of full ESP performance has been reached and proven in the past years. Despite this fact, further developments were carried out to keep partial functions of ESP available should failures occur, instead of switching off the system completely. Partial functions are, for example, that ESP is only available for free rolling and braking, or that ESP is only available during braking, or that basic ABS functions are only available, or that a controlled brake force distribution is

only given. The last item is of importance for vehicles with a critical brake force distribution. For instance, the system is switched to the first fallback stage—ESP only available for free rolling and braking—whenever a failure in the engine management interface occurs. In this case, full performance for all driving conditions except during drive, acceleration, and accelerator pedal lift-off is given. Almost the same holds true in case of a precharge pump failure (Figure 30.70).

In order to keep the basic ABS functionality available, even if one or more of the three sensors—steering wheel angle sensor, yaw rate sensor, or lateral acceleration sensor—fail, a new ABS reference speed algorithm without using one of these sensor signals was developed. It is designed to operate as a modular replacement for the standard velocity estimation of ESP and to work together with the unchanged BSC. This algorithm also uses the described adaptation phases to derive the free rolling wheel velocity. It is made robust by combining the reference speed of the wheels sidewise. The fallback algorithm runs in parallel to the standard velocity estimation of ESP. In the case of a sensor failure, the running ABS fallback reference speed module immediately replaces the standard one. At the same time, the vehicle motion controller is switched off and some state variables of the BSC are set to default values. In this manner, full ABS performance is achieved immediately after any failure of the above-mentioned three ESP sensors. Not only does the availability of the ABS function improve but so does the safety, since the system does not have to be switched off completely because of a sensor failure which occurs during active ESP control. As the BSC remains unchanged, and only the velocity estimation is replaced by the fallback reference speed module, the application of the fallback ABS control can be based on the application of the complete system. Especially the parameter values of the hydraulic model and most of the application parameter values are valid too for the fallback ABS.

The last fallback stage is required for vehicles with a critical brake force distribution in combination with the EBD in case all sensors including some (but not all) WSSs fail. This fallback stage will keep the EBD of the vehicle at an uncritical value by limiting the brake pressure at the rear axle (as long as the number of available WSSs will allow this).

30.4 COMPONENTS

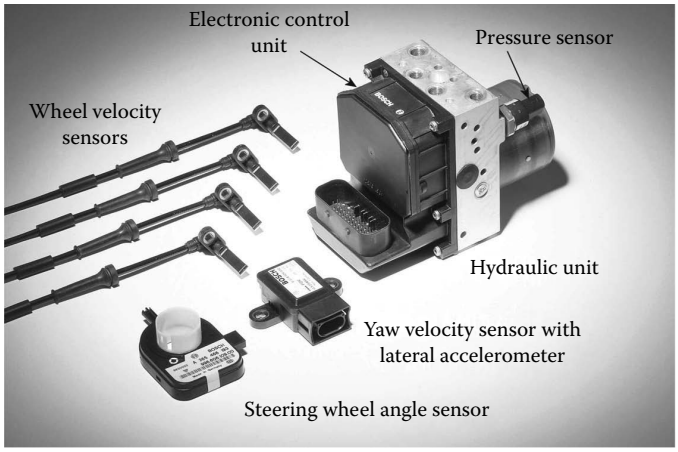
The components as chosen by the system definition are shown in Figure 30.64a, while Figure 30.64b shows them mounted in the car. The lateral accelerometer (i.e., the yaw rate sensor) should preferably be mounted close to the center of mass of the vehicle as this gives a good compromise in the phase shifts between the steering angle and the lateral acceleration at the front axle and that at the rear axle.

30.4.1 Sensors

30.4.1.1 Yaw Rate Sensor

The yaw rate sensor belongs to the class of vibrating gyros. Because of the importance of this sensor, there is much activity worldwide to develop low-cost precision sensors. The sensor used in the first ESP systems is described in [52]. The accuracy for the offset has to be better than ±2°/s and sensitivity errors have to be to less than ±5% over its life. It has a measuring range of ±100°/s, an externally toggled bipolar self-test function which causes a superimposed output of ±25°/s to the yaw velocity signal (Figure 30.59), and a reduced susceptibility to induced noise on the wiring and to radiated electromagnetic interferences. Figure 30.65 shows a low-cost micromachined sensor used as the successor of the first sensor mentioned earlier. For a detailed description of this sensor, see [53].

The yaw rate sensor is a twin sensor which combines two sensors. A micromachined yaw rate sensor element and a surface micromachined acceleration sensor element are integrated into one package. Its internal self-test capability switches, in case of a failure, the output into the out-of-range band of the signal so that internal defects can be detected by the ECU. The operating range of the acceleration signal is ±1.8 g. For improving the safety aspects, this design uses a background test for detecting internal defects too, which then switches the output in the out-of-range band of the signal.

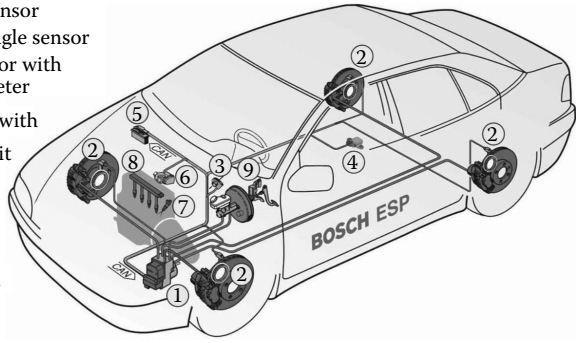


(a)

① Hydraulic unit with electronic control unit and pressure sensor
② Wheel velocity sensor
③ Steering wheel angle sensor
④ Yaw velocity sensor with lateral accelerometer

Engine management with

⑤ Electr. control unit
⑥ Throttle blade actuator
⑦ Injection valves
⑧ Ignition module
⑨ Accelerator pedal sensor



(b)

FIGURE 30.64 Components of the ESP system (a) and their placement in the car (b).

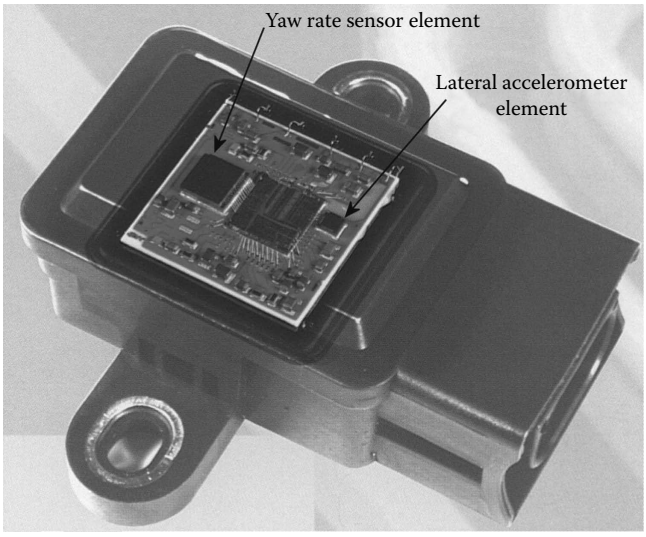


FIGURE 30.65 Yaw rate sensor with sensor hybrid shown on top.

Both outputs (yaw velocity and lateral acceleration) are analog and their values are obtained from the difference of the output signal with a +2.5 V reference voltage. The advantage of this method is that it improves the common mode rejection ratio of each signal path and therefore it reduces the susceptibility to various disturbances. A further step is the addition of a digital CAN interface which results in a further reduction of the amount of wiring and a further improvement in noise immunity.

The physical principle is the following (Figure 30.66). Two bulk-micromachined silicon masses are supported by leaf springs in a kind of a gimbal frame (side length 7 mm) and exited to oscillate by "Lorentz forces" at their natural frequency, typically of about 6 kHz. The oscillation of these two masses is in the antiphase mode and the mechanical coupling between these two masses by another pair of leaf springs is nearly one. The "gimbal" is supported by very weak springs to the environment so that the induced vibration from the vehicle is effectively isolated. On top of each of these masses, a surface micromachined accelerometer with open-loop signal processing is fixed to measure the Coriolis accelerations caused by the angular velocity of the sensor. The detection of internal defects can be signaled by switching the output into the out-of-range band of the signal. The offset stability is good and it is easy to compensate for temperature effects.

A similar accelerometer with closed-loop signal processing is used to measure the lateral acceleration of the vehicle (Figure 30.67) [53]. The accuracy for the offset has to be better than ± 0.075 g and sensitivity errors have to be less than ±7% over its life. A background test signal is continuously

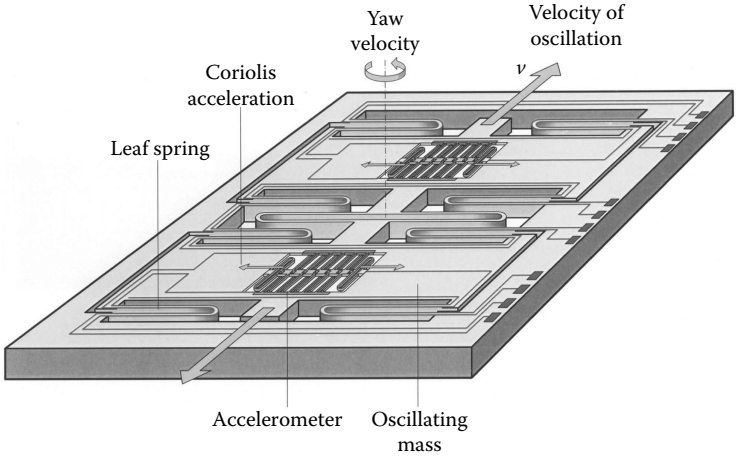


FIGURE 30.66 Yaw rate sensor element.

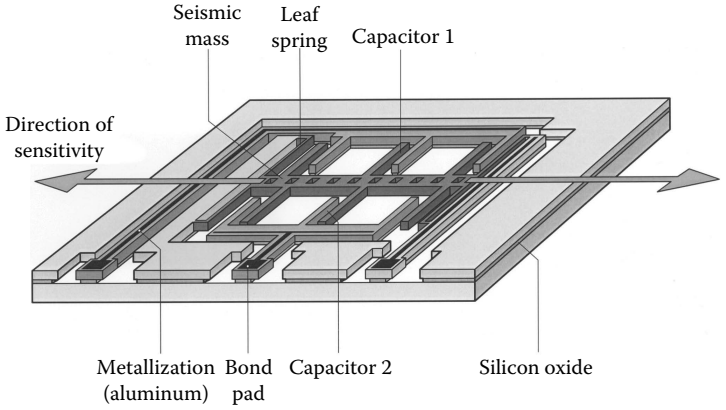


FIGURE 30.67 Accelerometer element used for the measurement of the Coriolis acceleration and the lateral acceleration.

fed into the signal-processing loop to detect internal defects and is then eliminated at the output to obtain a clean output signal. If this test signal differs by a certain amount, the output is switched into the out-of-range band of the signal.

The yaw rate sensor is supplied with nominal +12 V and therefore the outputs are not ratiometrical. Using a power supply regulator within the sensor improves the signal filtering also with respect to the increased power supply rejection ratio. ESP requires very high signal fidelity for these signals and therefore the common mode rejection ratio has to be large and the testability of these sensors becomes very important. This leads to a five-pin interface which is absolutely necessary to meet these high requirements. It allows reading the sensor output quasi-differential to compensate for transmission errors, to calibrate the A/D converter, and to enable the test mode. In total, signal path errors are reduced by a factor of two as compared to a ratiometrical three-wire interface. Digital interfaces using a CAN protocol are designed into the sensor. In this case, the susceptibility for conducted and radiated noise and corrosion and other problems with connectors is further reduced and the signal monitoring for inconsistent signal conditions is then much easier.

30.4.1.2 Steering Wheel Angle Sensor

The basic operating principle of the steering wheel angle sensor is the use of the anisotropic effect of a thin magnetic layer (AMR elements) [54]. Two of those elements are fixed to small gear wheels with different numbers of teeth and are driven by the steering column and their output signals are derived using a "nonius" principle to get a very high resolution. This special design also provides the absolute steering wheel angle position immediately after power up of the sensor so there is no need to have a sleep mode. Its operating range is ±720°. The overall accuracy has to be better than ±5° over its life. Figure 30.68 shows this sensor. The number of teeth of the small gear wheels is chosen in such a way that the difference in the rotation angle between the two can be used as a measure for the number of revolutions of the steering wheel. A microcontroller handles the four A/D converted output signals of the AMR elements, checks them if they are consistent, and converts them to the predefined CAN protocol. The susceptibility for conducted and radiated noise and corrosion and other problems with connectors is reduced and the signal monitoring for inconsistent signal conditions is then much easier. Its redundant signals and electronic paths and the implemented software concept allow a self-supervisory function on which the safety concept can rely. Internal defects are immediately sent via the CAN bus to the ECU. The CAN protocol has error detection and error correction capabilities. Therefore, the other sensors can be calibrated using this digital signal as a basis.

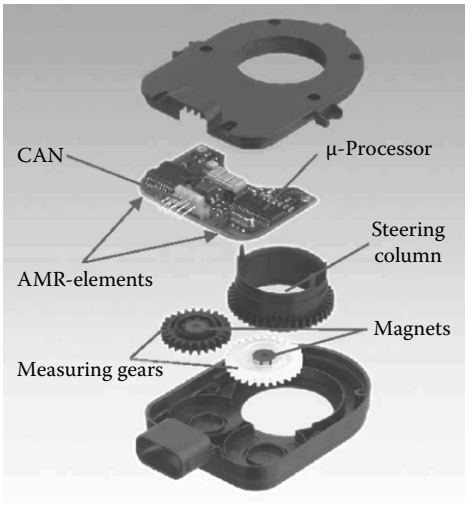


FIGURE 30.68 Exploded view of the steering wheel angle sensor.

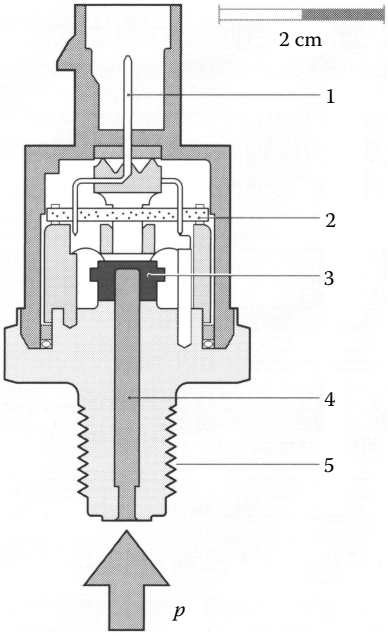


FIGURE 30.69 Cross section of the pressure sensor with (1) connector pins, (2) signal processing, (3) steel membrane with strain gages, (4) brake fluid bore, and (5) thread, p: brake fluid pressure.

30.4.1.3 Pressure Sensor

The pressure sensor has an operating range of 0–250 bar for measuring the pressure of the brake fluid either induced by the driver or by the precharge pump. The offset has to be less than ±3 bar and sensitivity errors have to be less than ±5% over its life. The sensor has been designed to include a background test which also allows the detection of internal defects that influence the offset and the sensitivity of the sensor. In case of an internal defect, the output of the sensor is switched into the out-of-range band of the signal. To ensure a correct function of this background test function, a power-up test sequence is initiated by the slew rate of the supply voltage at each power-up condition (Figure 30.60). This also allows detecting defects in the wiring harness and connectors between the sensor and the ECU.

The basic principle of this sensor is derived from the piezoresistive effect of a polysilicon layer designed as a strain gage, which is directly deposited on an electrically isolated membrane made of high-grade steel (Figure 30.69). The full bridge signal as the main signal path is amplified and compensated for temperature. The signal of one-half of the bridge is compared with a pressure-independent reference signal and the difference between these two signals is compensated for temperature and other effects and then amplified to show the same sensitivity as the signal of the main path. Thus, a test signal is derived to check the main signal for correctness in offset and sensitivity. If a deviation of a certain amount between these two signals occurs, the output is switched into the out-of-range band of the signal. This self-test function allows for the detection of sensitivity changes in the range of 5% or more and equivalent changes in the offset.

30.4.2 Actuators

30.4.2.1 Hydraulic Unit

The block diagram of a HU is shown in Figure 30.70. The unit is of the traction control type and is hydraulically closed using a self-priming recirculation pump (sRFP). For active pressure generation in the wheel brake cylinders, the active pressure valves, also called switch over valves, USV, are closed while the suction valves, HSV, are opened. The additional pressure-limiting function of the

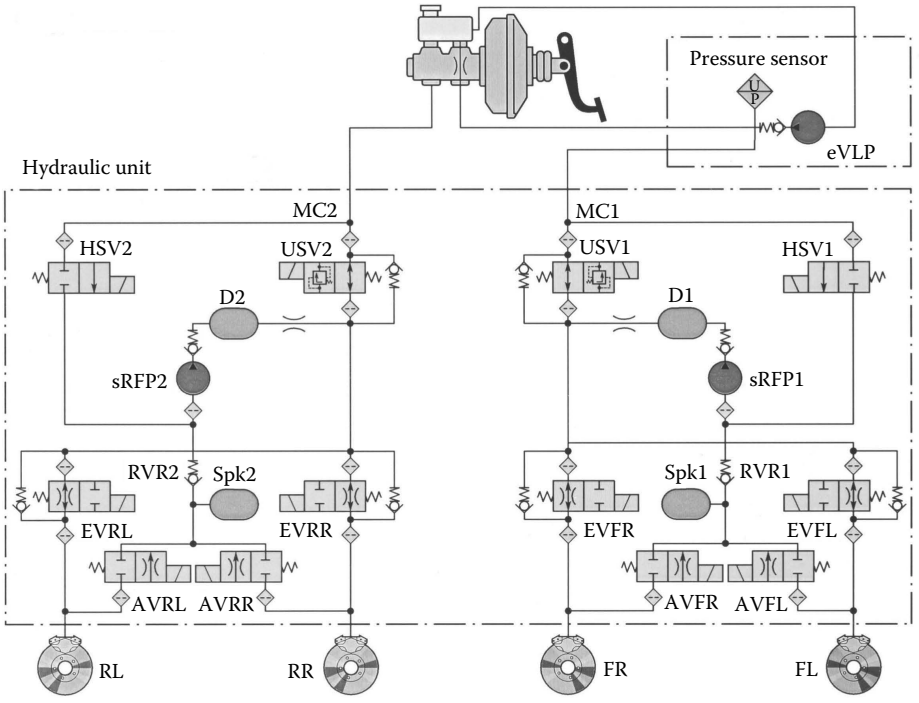


FIGURE 30.70 Hydraulic system of ESP with optional precharge pump.

USVs has a value which is chosen sufficiently high to reach locking pressure in each wheel brake cylinder at any maneuver. In order to limit the workload of the pump motor, the HSV is only opened if any of the EV in the corresponding brake circuit is open.

The suction valves ASV of the TCS–HU have to be replaced by the HSV in order to be able to open the valve also if the driver applies the brake. The HSV is a two-stage valve (Figure 30.71). In the first stage, a small seat in the valve is released. In the second stage, the larger seat is also released. If the driver applies the brake, the electromagnetic force of the HSV may not be strong enough for the anchor to release the larger seat also. However, during braking, it is sufficient to reach the first stage of the HSV only. For active braking, that is, if the driver does not apply the brake, the second stage must be opened. Since the driver does not apply the brake pedal, the electromagnetic force is large enough to release the large seat also. The design of the HSV allows it to release the small seat against MC pressures of up to 120 bar. This is sufficient, since at MC pressures above 120 bar all wheels are in the antilock braking mode.

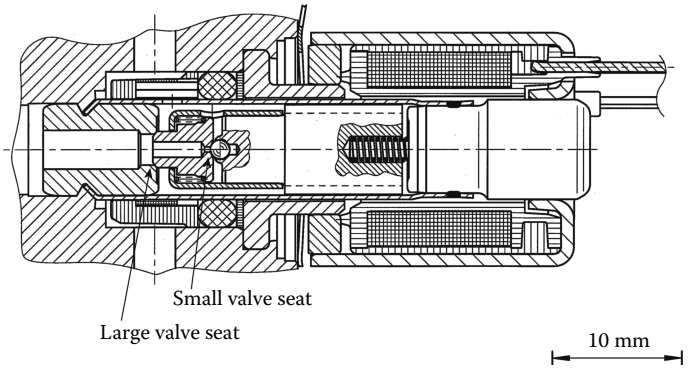


FIGURE 30.71 Cross section of the two-stage magnetic suction valve HSV.

30.4.2.2 Precharge Pump eVLP

The main reason for precharging the recirculation pump of the HU is the performance requirement of ESP at low ambient temperatures. Increasing viscosity of the brake fluid at lower temperatures together with the flow resistance of the brake fluid through the MC, the brake lines connecting the MC with the HU and the suction valves in the HU are the main reasons for the low efficiency of the sRFP. If a sRFP however, is capable of increasing the pressure in the wheel brakes fast enough by itself without precharging, then the precharge pump eVLP can be omitted. For most cars, precharging is not required. The precharge pump is of the gerotor type so that the fluid flow is almost continuous and there is no flow-restricting EV.

If the precharge pump is started, brake fluid flows from the reservoir through the precharge pump and the check valve into the push rod brake circuit. In the brake circuit, this fluid flow splits up and a part of the brake fluid flows through the brake MC back to the reservoir. The pressure in the brake circuit is determined by the fluid flow resistance from the brake MC back to the reservoir. This resistance is realized by a throttle bore in the MC whose diameter must be tuned together with the precharge pump delivery rate. The tuning must be such that the precharge pressure for the recirculation pump of the HU is sufficiently high (e.g., 15 bar) but not too high at low ambient pressures (otherwise the brake pedal is "hard" if the driver applies the brake during TCS operation). If the fluid flow passing through the MC back to the reservoir is reduced, then the brake circuit pressure is reduced as well. This occurs, for example, if the suction valve HSV is opened, and part of the fluid flow is led to the recirculation pump and eventually to the wheel brakes. Therefore, the precharge pump delivery rate must be clearly larger than that of the recirculation pump. Since the request for precharging is predominant at low temperatures, the precharge pump is designed for optimal delivery rate at low temperatures. During braking, the check valve prevents the brake fluid escaping from the push rod brake circuit through the precharge pump back to the reservoir.

Modern brake MCs have a floating piston which is not connected to the push rod piston. Forced by the pressure in the brake MC push rod circuit, the floating piston can therefore move away from its initial position. After the central valve of the floating piston has closed, the pressure in the floating piston brake circuit will increase to approximately the value of the push rod circuit if the floating piston friction is neglected. Thus, it is sufficient to connect the precharge pump to one brake circuit only. A failure in the precharge pump circuit (e.g., if the check valve leaks) will then not lead to a failure in the total brake system. Since the push rod circuit has a direct connection to the reservoir, pressure pulsations induced by the recirculation pump into the brake circuits are well damped.

30.4.3 Electronic Control Unit

The block diagram of the ECU is shown in Figure 30.72 and the nomenclature is shown at the end of this chapter. For safety reasons, the microprocessor is redundant similar to that with ABS and TCS and the cross-checking between the microprocessors is done as with the other systems (for cost reasons, modern systems use a different approach to check the proper function of the microprocessor). Also, critical signal paths are redundant to improve safety. For the voltage supply to the valves and to the pump motor, solid-state relays are used. The wheel velocity signals are output for use in other control units. The stabilized supply voltage of 5 V is required by, for example, the pressure sensor.

The ECU is mounted on the HU. This reduces cost, size, and wiring, and improves the reliability as compared with a separate ECU box. But since the HU is mounted with a bracket in the engine compartment, the ambient temperature of the ECU may be high (up to 120°C). The entire circuitry is therefore designed onto a ceramic substrate using the new small-scale hybrid technique. This substrate is directly attached to the HU. The coils of the valves are part of the ECU, so interconnection is very short and quite simple. This design results in small interferences

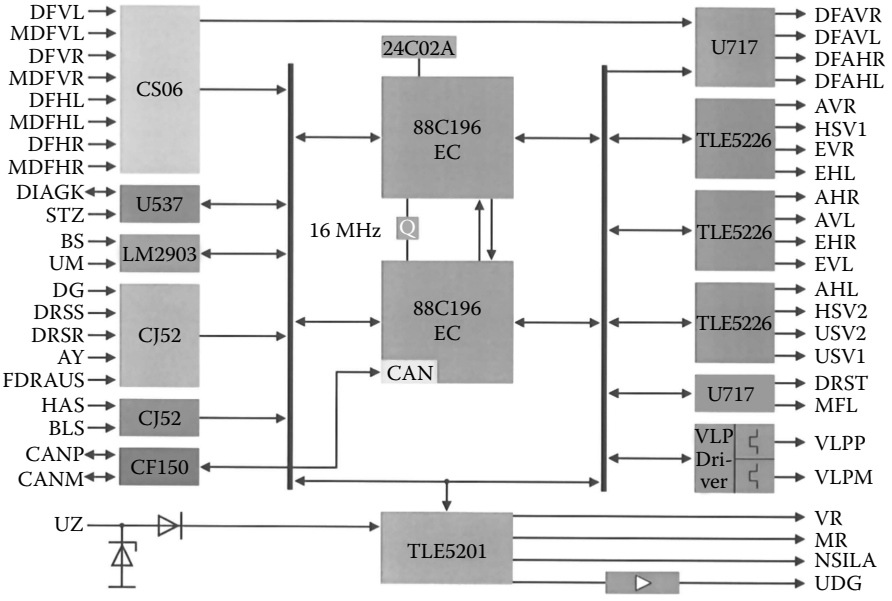


FIGURE 30.72 Block diagram of the ECU of ESP.

due to minor inductive coupling and therefore leads to an improved signal path performance and allows for pulse width-modulated stimulation of the valves and the pump motor. This results in the reduction of external wiring, which results in overall cost reduction. Each single signal path has been optimized concerning signal path errors and bandwidth aspect to achieve better overall accuracies for the sensor signals, which on the other hand also result in reduced sensor requirements.

A further step in reducing cost is achieved by the introduction of variant coding. Now it is possible to use the same ECU for different vehicles of the same platform and to choose the appropriate set of vehicle-dependent parameters like vehicle types, brake system, engine, and transmission. This results in fewer variants of ECUs and fewer spare parts in storage and also in more flexibility during the manufacturing process.

30.5 OUTLOOK

With the development of active four-wheel drive systems, active suspension systems, active rear wheel steering systems, and active brake and traction force distribution systems in the eighties of the last century, the desire for integration of these systems together with ABS and TCS followed [55–74]. In Europe, the goal was to improve the "vehicle safety to an unprecedented level" by using these systems in an integrated manner [65]. A summary is given in [74]. A drawback of this development was the immense cost and weight of those active systems to reach that level of safety. Only luxury cars could afford such high costs. With the introduction of ESP, the activities were virtually stopped since ESP already achieved a similar very high level of safety unknown up to that moment, but with significantly lower cost and with virtually no additional weight. Safety is thus not a privilege of luxury cars only but also of micro and compact cars.

New developments realize active front wheel steering systems and active antiroll bar systems which are not as expensive as the active systems of the eighties. Again the desire is to integrate these systems, however, now together with ESP [75–81]. Here, the approach of [76–78] is discussed.

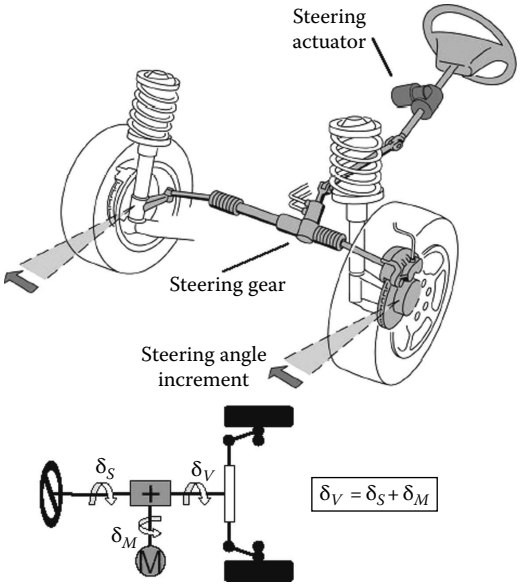


FIGURE 30.73 Active steering system.

If an active steering system is available, ESP can also use this system to control the handling behavior of the car, not only in limit situations but also during normal driving. Figure 30.73 shows the concept of a first realization of such a system with which ESP can modify the steering angle [82,84,85]. The system by itself performs several functions.

- Below a certain vehicle velocity, the steering transmission ratio is reduced to improve the handling of the vehicle and to improve maneuvering while parking in a slot.
- Compensation of the elasticity of the steering system and thus enhanced handling at higher velocities by feed forward control.
- Countersteering during split-μ braking, which reduces the steering load of the driver and which reduces the stopping distance because GMA (see ABS) is not required.
- Support of ESP by controlling the yaw moment of the vehicle through automatic steering. This feature is particular interesting since the steering angle corrections are very comfortable and the vehicle velocity is not influenced as when the yaw moment must be generated by brake and engine interventions. Moreover, the interventions are very fast so that the actuating signal can be kept small.

Another active system is shown in Figure 30.74, in which ESP can modify the normal force distribution on the wheels by control of special active antiroll bars as an active suspension [83,86]. The system itself performs several functions.

- Symmetric variation of the effective stiffnesses of the antiroll bars at the front and at the rear axle to reduce the roll angle while cornering.
- Asymmetric variation of the effective stiffnesses of the antiroll bars at the front and at the rear axle to influence the handling behavior of the vehicle in the direction of understeer or oversteer.
- Asymmetric variation of the antiroll bars at the front and at the rear axle to increase the normal, lateral, and longitudinal forces on the wheels and enhance the stopping distance during split-μ braking or reduce the yaw moment on the vehicle.

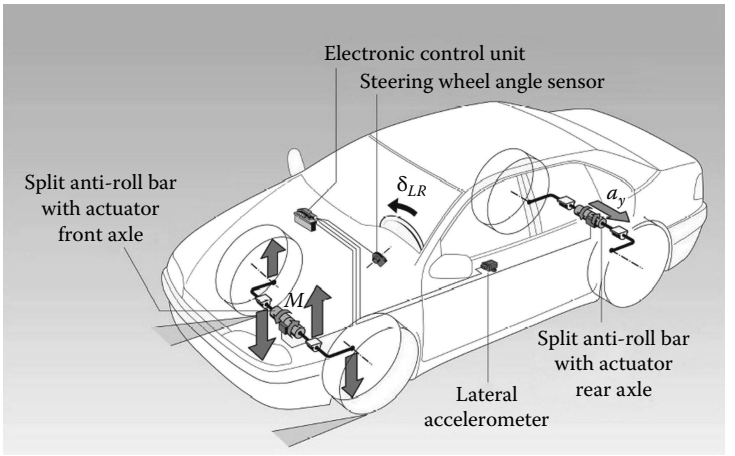


FIGURE 30.74 Active antiroll bar.

- Support of ESP by controlling the yaw moment of the vehicle through automatic variation of the effective stiffnesses of the antiroll bars on the front and on the rear axle. This feature is particular interesting since the antiroll bar corrections are, like the steering angle corrections, very comfortable and the vehicle velocity is not influenced as when the yaw moment must be generated by brake and engine interventions. However, in contrast with the steering angle corrections, the interventions by the active antiroll bars are relatively slow so that the actuating signal cannot be kept small by this feature alone.

Each of these systems can thus influence the yaw moment on the car. The question is now, how to implement the total yaw rate control (the term yaw rate control means the vehicle dynamics control). Early proposals were based on the observation of the different intervention bandwidths of the brake, engine, active steering, and active suspension actuators. Each actuator had its own yaw rate controller (Figure 30.75). The controller gains had to be tuned such that the interference between the interventions did not lead to undesirable or even unstable vehicle behavior. The charm of this approach is that to some extent the development of the controllers can be done independent of each other. Obviously, if the controller gains must be drastically reduced

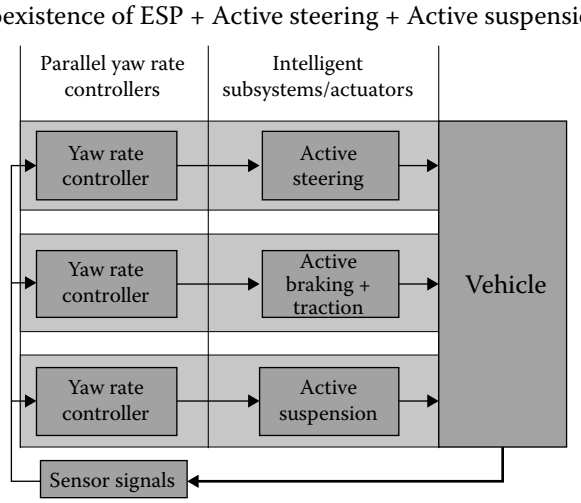


FIGURE 30.75 System integration using parallel yaw rate controllers.

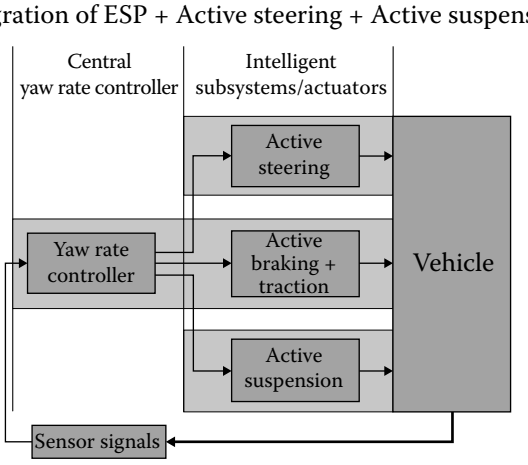


FIGURE 30.76 System integration using a central yaw rate controller.

to guarantee the peaceful coexistence of the controllers, important performance potential is lost. Identification, adaptation, and learning control are not feasible. Furthermore, the systems must exchange information. This can be seen from an intervention of the active suspension. If the active suspension modifies the behavior of the vehicle and changes the characteristic velocity, then ESP may intervene and compensate the modification in the vehicle behavior. The complexity of the control and the controller safety will increase rapidly with the number of systems and so does the development and application effort. The question of multiple sensors or inclusion of nonstandard sensors in the control and safety concepts must be solved also.

Later proposals use a central yaw rate controller which takes the different properties of the actuators in consideration (Figure 30.76). This is particularly important if the brake actuation is no more much slower than the steering actuation as is the case with an EHB where the brake interventions can be very fast. The superimposed central yaw rate control distributes the required yaw moment between the active systems and uses them as intelligent actuators. The idea is to extend the structure of the existing ESP in which the "intelligent subsystems" BSC and TSC determine their own target values of the controlled variable "slip" and where the vehicle dynamics controller determines the increments of the slip to adjust the yaw moment on the vehicle. Thus, the active steering controller and the active suspension controller must determine their target values by themselves while the vehicle dynamics controller is extended to also determine the increments of the steering angle and of the effective antiroll bar stiffnesses to adjust the yaw moment on the vehicle.

EHB and the new active systems can influence the handling behavior of the car also in situations in which the physical limit is not reached, for example, by a lateral brake force distribution in the case of EHB. If the passive car handling changes, for example, because of changes in the tire properties, then these systems can restore normal vehicle behavior by interventions which are not noticed by the driver. These systems therefore have the potential of making handling robust against parameter changes. However, in order to fully exploit the high speed of the interventions, the test cycle of the yaw velocity signal (Figure 30.59) must be avoided, which results in the requirement for redundant yaw velocity signals.

NOMENCLATURE

$\dot{\psi}$	yaw velocity
$\dot{x}_0$	velocity of the vehicle at the center of mass in x_0-direction
$\dot{y}_0$	velocity of the vehicle at the center of mass in y_0-direction
$\ddot{x}_0$	acceleration of the vehicle at the center of mass in x_0-direction
$\ddot{y}_0$	acceleration of the vehicle at the center of mass in y_0-direction

$\tilde{v}_y$	velocity of the vehicle at the center of mass in the lateral direction as obtained from the measurement equation
$\hat{v}_y$	velocity of the vehicle at the center of mass in the lateral direction as obtained from the Kalman filter equation
$\dot{\beta}$	slip angle velocity of the vehicle at the center of mass
$\dot{\psi}_M$	yaw velocity of the extended bicycle model
$\dot{v}_y$	time rate of change of the lateral velocity of the vehicle at the center of mass
$\dot{\dot{\psi}}$	yaw acceleration
$\dot{v}_{Whl}$	acceleration of the wheel
$\Delta\dot{\psi}$	difference between the nominal and the actual yaw velocity values
$\dot{v}_x$	time rate of change of the longitudinal velocity of the vehicle at the center of mass
$\dot{v}_{x,off}$	slope of the road in longitudinal direction of the vehicle
$\dot{v}_{x,off}$	time rate of change of the slope of the road in longitudinal direction of the vehicle
$\dot{\psi}_M$	yaw velocity of the extended bicycle model of the vehicle
$\Delta\dot{\psi}_{No}$	actuating signal of the yaw velocity of the vehicle
$\Delta\dot{\psi}_b$	compensation of the yaw velocity of the extended bicycle model for banked turns
$\dot{\psi}_{No}$	nominal yaw velocity
$\dot{\psi}_k$, $\dot{\psi}_{k+1}$	yaw velocity at sample time point k and $k + 1$ respectively
$+A$	acceleration threshold
$+a$	acceleration threshold
a	longitudinal distance from the front axle to the center of mass of the vehicle
A	bulk head area of the vehicle
$-a$	deceleration threshold
A_0, A_1, A_2	coefficients for the computation of the target wheel slip
A_{11}, A_{12}, A_{21}, A_{22}	time variable coefficients
a_{Whl}	acceleration of the wheel
a_x, a_y	acceleration in longitudinal resp. lateral direction of the vehicle at the center of mass
ay_{corr}	calibrated lateral acceleration signal
ay_{Mess}	measured lateral acceleration signal
ay_{off}	offset value of the measured lateral acceleration signal
ay_{ref}	lateral acceleration value computed from the reference value of the yaw velocity
b	half the track width
c	longitudinal distance from the rear axle to the center of mass of the vehicle
C_p	brake torque ratio
c_w	wind drag coefficient
C_α, C_λ	lateral resp. longitudinal tire stiffness
$C_{\alpha,F}$, $C_{\alpha,R}$	total lateral stiffness at the front and rear wheels respectively
D_T	sampling time of the brake slip controller
D_λ	dead zone for the drive slip difference between the driven wheels of one axle
$e_\Delta v$	actuating signal for the drive slip difference controller
F_1, …, F_4	lateral force from the wheel on the vehicle at wheel 1, …, 4
F_B	brake force from the road on the tire
F_b	braking force on the wheel at the brake pad
F_b^*	traction force increment due to braking another wheel
$F_{B,A}$	brake force from the road on the tire during the adaptation phase
F_{B1}, …, F_{B4}	brake force from the road on the tire for wheel number 1, …, 4
F_{BF}	stationary (filtered) brake force from the road on the tire
F_h	traction force from the road on the driven wheel at the high-friction road surface
F_l	traction force from the road on the driven wheel at the low-friction road surface
F_N	normal force from the road on the tire
F_R	resultant force from the road on the tire in the plane of the road

F_{R1}	resultant force from the road on the tire of wheel number 1 in the plane of the road
F_S	lateral force from the road on the tire
$F_{S,F}$	lateral force from the road on the tire at the front wheel of the bicycle model
$F_{S,R}$	lateral force from the road on the tire at the rear wheel of the bicycle model
F_{S1}, …, F_{S4}	side force from the road on the tire at wheel number 1, …, 4
F_x	longitudinal tire force
F_y	lateral tire force
f_ω	sensitivity change in the measured yaw velocity signal
I_{EVi}, I_{AVi}	currents for the inlet resp. outlet valves of the EHB system
I_{TR}	total transmission ratio of the drive and power trains
J_{Mot}	engine crankshaft moment of inertia
J_V	moment of inertia of the vehicle about the body fixed z-axis
J_{Whl}	moment of inertia of the wheel about its axis
K_p, K_d, K_i	gains for the PID brake slip controller
Lw_{corr}	calibrated steering wheel angle signal
Lw_{Mess}	measured steering wheel angle signal
Lw_{off}	offset value of the measured steering wheel angle signal
Lw_{ref}	steering wheel angle value computed from the reference value of the yaw velocity
m	lowest negative nominal torque on the driven wheels for MSR control
M_1	oversteering yaw moment
M_2	understeering yaw moment
M_{BR}	brake torque from the brake on the wheel
$M_{BR,A}$	brake torque from the brake on the wheel during the adaptation phase
M_{Ca}	torque of the cardan axle multiplied by the transmission ratio of the differential
M_{CaHalf}	drive torque on a driven wheel
M_{Dif}	difference in the brake torques between two driven wheels at one axle
M_{Dr}	output torque of the engine requested by the driver
M_{MOT}	engine output torque
M_{NoLock}	nominal brake torque difference for the control of the difference between the drive slips of the driven wheels of one axle
M_{NoMot}	nominal engine output torque for throttle control
M_{NoSPR}	nominal engine torque reduction by spark retard
M_R	torque from the road on the wheel
m_v	vehicle mass
M_{WhlNo}	nominal torque on the wheel
M_{WhlNo3}, M_{WhlNo4}	nominal torque on driven wheels 3 and 4, respectively
M_{YawNo}	nominal yaw moment
M_{Yw}, M_{yaw}	yaw moment
p_{Circ}	pressure in the brake circuit
p_{Whl}	pressure in the brake wheel cylinder
$p_{Whl,i}$	pressure in the brake wheel cylinder number i
p_{WhlNo}	nominal pressure in the brake wheel cylinder
p_{WhlNoi}	nominal pressure in brake wheel cylinder number i
R	wheel radius
R_T	radius of the turn
SUM	event-controlled integration
T	time length between adjacent sampling time points
t_1, t_2	start resp. end time of the increase in the pressure difference threshold between two brake wheel cylinders at one axle
t_{EVi}, t_{AVi}	stimulation times of the inlet resp. outlet valve of wheel number i
T_{iOFF}	time period during which the fuel injection is shut off of engine cylinder number i

u_1, u_2	forcing functions
$u_{1,k}$, $u_{2,k}$	forcing functions at sample time point k
U_{Mval}	stimulation time length of the inlet and outlet electromagnetic valves
v_{Ca}	average velocity of the driven wheels of one axle
v_{CH}	characteristic velocity
v_{Dif}	difference in the velocities between two driven wheels at one axle
v_i	calibrated wheel velocity signal
$v_{i,Mess}$	measured wheel velocity signal
v_{NoCa}	nominal average velocity of the driven wheels of one axle
v_{NoDif}	difference in the nominal wheel velocities of the driven wheels of one axle
v_{Ref}	reference velocity
v_V	velocity of the vehicle at the center of mass
v_{VWhl}	velocity of the wheel center
$v_{VWhl,F}$	velocity of the front wheel center of the bicycle model
$v_{VWhl,R}$	velocity of the rear wheel center of the bicycle model
v_{Whl}	velocity of the wheel
$v_{Whl,1}$, $v_{Whl,4}$	velocity of wheel number 1, …, 4
$v_{Whl,A}$	wheel velocity during the adaptation phase
v_{WhlFre}	free rolling wheel velocity
$v_{WhlFre3}$,$v_{WhlFre4}$	free rolling velocity of wheel number 3 resp. 4
v_{WhlNo3},v_{WhlNo4}	nominal velocity of wheel number 3 resp. 4
v_x	velocity of the vehicle in longitudinal direction at the center of mass
v_y	velocity of the vehicle in lateral direction at the center of mass
$v_{y,k}$, $v_{y,k+1}$	velocity of the vehicle center of mass in lateral direction at sample time points k and $k + 1$, respectively
x_1, x_2	parameters for the brake hydraulics simulation model
α	slip angle of the wheel
α_0	slip angle at the maximum of the μ-slip angle curve
α_F	slip angle of the front wheels
α_R	slip angle of the rear wheels
β	slip angle of the vehicle at the center of mass
$\beta\,(t)$	time history of the slip angle of the vehicle at the center of mass
β_0	initial value of the slip angle of the vehicle at the center of mass
β_M	slip angle of the extended bicycle model of the vehicle at the center of mass
ΔM_{NoLock}	reduction in the nominal brake torque difference for the control of the difference between the drive slips of the driven wheels of one axle determined by the vehicle dynamics controller
δ, δ_S, δ_{LR}	steering wheel angle
δ_A	Ackermann steering angle
Δa_{yb}	compensation of the lateral acceleration for banked turns
ΔD_λ	increment in the dead zone for the drive slip difference between the driven wheels of one axle determined by the vehicle dynamics controller
δ_M	steering wheel angle increment from active steering
ΔM_{Yw}	increment in the yaw moment
Δp_1	initial threshold in the pressure difference between two brake wheel cylinders at one axle
Δp_2	final threshold in the pressure difference between two brake wheel cylinders at one axle
δ_V	total steering wheel angle for active steering
δ_W	steering angle of the front wheels
$\Delta\beta$	difference between the nominal and the actual vehicle slip angle values
$\Delta\beta_{No}$	actuating signal of the slip angle of the vehicle
$\Delta\lambda$	increment in the wheel slip computed by the vehicle dynamics controller

$\Delta\lambda_1$	increment in the slip of wheel number 1
$\Delta\lambda_i$	increment in the slip at wheel number *i*
$\Delta\lambda_{MA}$	increment in the nominal average drive slip from the vehicle dynamics controller
$\Delta\lambda_T$	increment in the target slip for off-road conditions
$\Delta\omega$	difference between the calibrated and reference yaw velocity values
$\Delta\omega_{zul}$	failure threshold of the yaw velocity fault
λ	wheel slip
λ_0	slip value at the maximum of the μ-slip curve
λ_1	Slip threshold velocity
λ_A	slip during the adaptation phase
λ_i	slip of wheel number *i*
λ_{MA}	nominal average drive slip of the driven wheels of one axle
λ_{MAmin}	average drive slip of the driven wheels of one axle with maximum vehicle stability during the TCS function
λ_{MAT}	target average drive slip of the driven wheels of one axle
λ_{No}	nominal wheel slip
λ_{Noi}	nominal slip of wheel number *i*
λ_T	target slip
λ_{Ti}	target slip of wheel number *i*
λ_{WhlNo3}, λ_{WhlNo4}	nominal slip of wheel number 3 resp. 4
μ_A	coefficient of friction in longitudinal wheel direction during the adap tation phase between the tire and the road surface
μ_B	brake coefficient of friction between the tire and the road surface
μ_{BM}	coefficient of friction at the maximum of the μ-slip curve
μ_{BS}	coefficient of friction between the tire and the road of the locked wheel
μ_h	high value of the coefficient of friction of the road surface
μ_l	low value of the coefficient of friction of the road surface
μ_{Res}	coefficient of friction of the road surface computed from the longitudinal and lateral forces on the wheel
μ_S	lateral coefficient of friction between the tire and the road surface
ρ	mass density of air
ψ	yaw angle
ω_{ay}	yaw velocity value computed from the lateral acceleration value
ω_{corr}	calibrated yaw velocity value
ω_{Lw}	yaw velocity value computed from the steering wheel angle value
ω_{Mess}	uncalibrated yaw velocity value
ω_{off}	offset value of the measured yaw velocity signal
ω_R	rotational wheel velocity
ω_{R0}	free rolling rotational wheel velocity
ω_{ref}	reference value of the yaw velocity
ω_{streu}	influence of faults on the signals of the wheel velocities, lateral accel eration, and steering wheel angle on the fault in the reference value of the yaw velocity
ω_v	yaw velocity value computed from the wheel velocity values

ACRONYMS

ABS	antilock brake system
AC	alternating current
A/D	analog to digital converter
AMR	anisotropic magneto resistive
ASR	traction control system

ASV	suction valve
AV	outlet valve
AVHL	outlet valve rear left
AVHR	outlet valve rear right
AVVL	outlet valve front left
AVVR	outlet valve front right
AY	lateral acceleration sensor input
B	brake booster
BA	brake assistant
BC	brake wheel cylinder
BITE	build in test equipment
BLS	brake light switch
BS	brake switch
BSC	brake slip controller
CAN	controller area network
CANM	CAN bus ground
CANP	CAN bus signal
D1	damping chamber front
D2	damping chamber rear
DFAHL	wheel speed sensor output rear left
DFAHR	wheel speed sensor output rear right
DFAVL	wheel speed sensor output front left
DFAVR	wheel speed sensor output front right
DFHL	wheel speed sensor rear left
DFHR	wheel speed sensor rear right
DFVL	wheel speed sensor front left
DFVR	wheel speed sensor front right
DG	pressure sensor input
DIAGK	diagnostics input/output
DRSR	yaw rate sensor reference voltage
DRSS	yaw rate sensor signal
DRST	yaw rate sensor test signal
EBD	electronic brake force distribution
ECU	electronic control unit
EEPROM	electrically erasable programmable read only memory
EHB	electro hydraulic brake system
EM	electric motor for recirculation pump
EMS	electronic engine power control
ESP	electronic stability program
EV	inlet valve
EVHL	inlet valve rear left
EVHR	inlet valve rear right
eVLP	precharge pump new generation
EVVL	inlet valve front left
EVVR	inlet valve front right
FDR	electronic stability program
FDRAUS	ESP-ON/OFF-switch
FMEA	failure mode and effect analysis
FWD	front wheel drive
GMA	yaw moment gradient attenuation
HAS	hand brake switch

HBA	hydraulic brake assistant
HDC	hill descent control
HSV	high pressure suction valve
HU	hydraulic unit
M	electric motor
MC	brake master cylinder
MDFHL	wheel speed sensor rear left ground
MDFHR	wheel speed sensor rear right ground
MDFVL	wheel speed sensor front left ground
MDFVR	wheel speed sensor front right ground
MFL	multi function lamp
MR	pump motor relay
MSR	engine drag torque control
NSILA	SILA (inverted)
PATA	ESP-ON/OFF-signal input
PBA	pneumatic brake assistant
PCB	printed circuit board
PI	proportional-integral-type controller
PID	proportional-integral-differential-type controller
RAM	random access memory
RFP	recirculation pump
ROM	read only memory
ROM	roll over mitigation
RTA	wheel velocity fault compensation
RVR1	check valve front
RVR2	check valve rear
RWD	rear wheel drive
SL	select low
SN	brake fluid level signal input
Spk1	accumulator front
Spk2	accumulator rear
sRFP	self-priming recirculation pump
STZ	idle timer
SW	software
TCS	traction control system
TSC	traction slip controller
UDG	power supply of pressure sensor
UM	voltage from recirculation pump
USV	pressure limiting electromagnetic valve
UZ	power supply input
VDA	association of german vehicle manufacturers
VLPM	precharge pump ground
VLPP	precharge pump voltage
VR	valve relay
WSS	wheel speed sensor

REFERENCES

1. *Fahrsicherheitssysteme*, Bosch, Vieweg, Wiesbaden, Hesse, Germany, 1998.
2. Bert Breuer, Karl-Heinz Bill (eds.) *Bremsenhandbuch,* 4th edition, Springer Vieweg, Wiesbaden, Germany, 2013.
3. ABS/TCS/VDC Technology Report 1996. *Tier One*, Sacramento, CA, 1996.

4. Furukawa, Y., Abe, M., Advanced chassis control systems for vehicle handling and active safety. *Vehicle System Dynamics*, 28, 59–86, 1997.
5. Inagaki, S., Kshiro, I., Yamamoto, M., Analysis on vehicle stability in critical cornering using phase-plane method. *AVEC'94*, Nr. 9438411, Tsukuba, Japan, 1994, pp. 287–292.
6. Leiber, H., Czinczel, A., Der elektronische Bremsregler und seine Problematik. *ATZ Automobiltechnische Zeitschrift*, 74(7), 269–277, 1972.
7. Maisch, W., Jonner, W., Sigl, A., ASR—Traction control—A logical extension of ABS. *SAE'87*, Nr. 870337, Detroit, USA, 1987.
8. Egger, G., Herb, E., Krusche, H., Wallentowitz, H., Stabilitäts- und Traktionsüberwachung bei allradgetriebenen Personenwagen. *ATZ Automobiltechnische Zeitschrift*, 91(2), 83–90, 1989.
9. Shiraishi, S., Yamamoto, O., Kin, K., Akuta, Y., Traction control for improved vehicle dynamics. *JSAE Review*, 13(1), 58–64, 1992.
10. Böning, B., Folke, R., Franzke, K., Traction control (ASR) using fuel-injection suppression—A cost effective method of engine-torque control. *SAE 92*, Nr. 920641, Detroit, USA, 1992.
11. van Zanten, A., Krauter, A., Optimal control of the tractor-semitrailer truck. *Vehicle System Dynamics*, 7, 203–231, 1978.
12. Zomotor, A., Leiber, H., Schlupfregelung—Ein Weg zur Erhöhung der aktiven Sicherheit. *Fachtagung der VDI-Gesellschaft Fahrzeugtechnik Reifen Fahrwerk Fahrbahn*, Hannover, Germany, September 24–25, 1987.
13. Heeß, G., van Zanten, A. T., System approach to vehicle dynamics control, *Fisita 1988*, Nr. 885107, Detroit, MI, 1988, pp. 2.109–2.121.
14. McLellan, D., Ryan, J., Browalski, E., Heinricy, J., Increasing the safe driving envelope—ABS, Traction control and beyond. *SAE 92C014*, Huntsville, AL, 1992.
15. Sato, S., Inoue, H., Tabata, M., Inagaki, S., Integrated chassis control system for improved vehicle dynamics. *AVEC'92*, Nr. 923074, Yokohama, Japan, 1992, pp. 413–419.
16. Matsumoto, S, Yamaguchi, H., Inoue, H., Yasuno, Y., Braking force distribution control for improved vehicle dynamics. *AVEC'92*, Nr. 923079, Yokohama, Japan, 1992, pp. 441–446.
17. Shibahata, Y., Shimada, K., Tomari, T., The improvement of vehicle maneuverability by direct yaw moment control. *AVEC'92*, Nr. 923081, Yokohama, Japan, 1992, pp. 452–457.
18. Motoyama, S., Uki, H., Isoda, K., Yuasa, H., Effect of traction force distribution control on vehicle dynamics. *AVEC'92*, Nr. 923080, Yokohama, Japan, 1992, pp. 447–451.
19. Müller, A., Achenbach, W., Schindler, E., Wohland, T., Mohn, F.-W., Das neue Fahrsicherheits system electronic stability program von Mercedes Benz. *ATZ Automobiltechnische Zeitschrift*, 96(11), 656–670, 1994.
20. van Zanten, A., Erhardt, R., Pfaff, G., VDC, The vehicle dynamics control system of Bosch, *SAE 95*, Nr. 950759, Warrendale, PA, 1995.
21. Koibuchi, K., Yamamoto, M., Fukada, Y., Inagaki, S., Vehicle stability control in limit cornering by active brake. *SAE 96*, Nr. 960487, Aachen, Germany, 1996.
22. Yasui, Y., Tozu, K., Hattori, N., Sugisawa, M., Improvement of vehicle directional stability for transient steering maneuvers using active brake control. *SAE 96*, Nr. 960485, Detroit, USA, 1996.
23. Gass, H., Glaser, H., Heißing, B., Mäusbacher, B., Electronic driving aids. *AVEC'96, Proceedings International Symposium on Advanced Vehicle Control*, Aachen, Germany, June 24–28, 1996, pp. 609–627.
24. Alberti, V., Babbel, E., Improved driving stability by active braking of the individual wheels. *AVEC'96, Proceedings International Symposium on Advanced Vehicle Control*, Aachen, Germany, June 24–28, 1996, pp. 717–732.
25. Jost, K., Cadillac, stability enhancement, *Automotive Engineering*, pp. 111–113, October 1996.
26. Debes, M., Herb, E., Müller, R., Sokoll, G., Straub, A., Dynamische stabilitäts control DSC der Baureihe 7 von BMW—Teil 2. *ATZ Automobiltechnische Zeitschrift*, 99(4), 208–213, 1997.
27. Sugiyama, M., Inoue, H., Uchida, K., Monzaki, S., Inagaki, S., Kido, S., Development of VSC (vehicle stability control) system. *Toyota Technical Review*, 46(2), April 1997.
28. Fennel, H., ABS plus und ESP—Ein Konzept zur Beherrschung der Fahrdynamik. *ATZ Automobiltechnische Zeitschrift*, 100(4), 302–308, 1998.
29. Fennel, H., Gutwein, R., Kohl, A., Latarnik, M., Roll, G., Das modulare Regler- und Regelkonzept beim ESP von ITT Automotive, 7. *Aachener Kolloquium Fahrzeug- und Motortechnik*, Aachen, Germany, Oktober 5–7, 1998, S409–S431.
30. Straub, T., Vehicle stability control (VSC) and its rapid product improvement cycle to meet a wide customer base. *TOPTEC,* Wien, Austria, 1999.
31. Sawase, K., Sano, Y., Application of active yaw control to vehicle dynamics by utilizing driving/braking force. *JSAE 9930801*, 1999.

32. van Zanten, A., Erhardt, R., Landesfeind, K., Pfaff, G., VDC systems development and perspective. *SAE 98*, Nr. 980235, 1998.
33. van Zanten, A., Bosch ESP systems: 5 years of experience. *SAE Automotive Dynamics Stability Conference*, Paper Nr. 2000–01–1633, Troy, MI, May 15–17, 2000.
34. Hattori, Y., Koibuchi, K., Yokoyama, T., Force and moment control with nonlinear optimum distribution for vehicle dynamics. *AVEC'02, JSAE* 20024577, Yokohama, Japan, 2002.
35. Förster, H.-J., Der Fahrzeugführer als Bindeglied zwischen Reifen, Fahrwerk und Fahrbahn. *VDI Berichte*, Nr. 916, Düsseldorf, Germany, 1991.
36. van Zanten, A., Erhardt, R., Lutz, A., Measurement and simulation of transients in longitudinal and lateral tire forces. *SAE 90,* Nr. 900210, Detroit, USA, 1990.
37. Gottwick, U., Schunck, E., van Zanten, A., Die Elektrohydraulische Bremse revolutioniert die Pkw-Bremssysteme. *VDI Berichte*, Nr. 1496, Düsseldorf, Germany, 1999, pp. 287–303.
38. Tseng, H., Dynamic estimation of road bank angle. *Vehicle System Dynamics*, 36(4–5), 307–328, 2001.
39. van Zanten, A., Erhardt, R., Pfaff, G., Kost, F., Hartmann, U., Ehret, T., Control aspects of the Bosch-VDC. *AVEC'96*, *Proceedings International Symposium on Advanced Vehicle Control*, Aachen, Germany, June 24–28, 1996, pp. 574–607.
40. Fukada, Y., Slip-angle estimation for vehicle stability control. *Vehicle System Dynamics*, 32, 375–388, 1999.
41. Furukawa, Y., Abe, M., Direct yaw moment control with estimating side-slip angle by using on-board-tire-model. *AVEC'98*, *JSAE* 9837030, 1998.
42. Venhovens, P., Naab, K., Vehicle dynamics estimation using kalman filters. *AVEC'98*, *JSAE* 9836617, 1998.
43. Kato, M., Isoda, K., Yuasa, H., Estimation of vehicle side slip angle with artificial neural network. *JSAE Review*, 15, 79–82, *JSAE* 9430103, 1994.
44. Meditch, J., *Stochastic Optimal Linear Estimation and Control*. McGraw Hill, New York, 1969.
45. van Zanten, A., Erhardt, R., Lutz, A., Neuwald, W., Bartels, H., Simulation for the development of the Bosch—VDC. *SAE 96*, Nr. 960486, Detroit, USA, 1996.
46. Fischer, G., Müller, R., Das elektronische Bremsenmanagement des BMW X5. *ATZ Automobiltechnische Zeitschrift*, 102(9), 764–773, 2000.
47. Ponticel, P., Dynamic rollover testing on the way. *Automotive Engineering,* pp. 26–28, November 2003.
48. Wielgat, A., SUV Safety: The Volvo Way. *Automotive Industries*, pp. 36–38, June 2002.
49. Fischer, G., Heyken, R., Trächtler, A., Aktive Gespannstabilisierung beim BMW X5, *ATZ Automobiltechnische Zeitschrift*, 4/2002, Jahrgang 104, 2002, pp. 330–336.
50. Kiesewetter, W., Klinkner, W., Reichelt, W., Steiner, M., Der neue Brake Assist von Mercedes-Benz—Aktive Fahrerunterstützung in Notsituationen. *ATZ Automobiltechnische Zeitschrift*, 99(6), 330–339, 1997.
51. Lücke, H., Wagner, J., Wiss, H., Hydraulischer Bremsassistent. *ATZ, Automobiltechnische Zeitschrift*, 101(6), 470–475, 1997.
52. Aanen, G., Plausibilitätsbasierte Fehlererkennung. MSc thesis, Number A-740, Technical University Delft, Delft, the Netherlands, 1996.
53. Fennel, H., Ding, E., A model-based failsafe system for the continental TEVES electronic-stability-program (ESP). *SAE Automotive Dynamics Stability Conference*, Paper Nr. 2000–01–1635, Troy, MI, May 15–17, 2000, pp. 229–237.
54. Isermann, R., Diagnosis methods for electronic controlled vehicles. *Vehicle System Dynamics*, 36(2–3), 77–117, 2001.
55. Reppich, A., Willig, W., Yaw rate sensor for vehicle dynamics control system, *SAE 95*, Nr. 950537, Detroit, USA, 1995.
56. Lutz, M., Golderer, W., Gerstenmeier, J., Marek, J., Maihöfer, B., Schubert, D., A precision yaw rate sensor in silicon micromachining. *SAE 98,* Nr. 980267, Dearborn, MI, 1998.
57. Gruber, J., Steering wheel angle sensor for vehicle dynamics control systems. *SAE 97*, Nr. 970382, St. Louis, MO, 1997.
58. Shibahata, Y., Irie, N., Ito, H., Nakamura, K., The development of an experimental four-wheel-steering vehicle. *SAE 86*, Nr. 860623, Detroit, USA, 1986.
59. Kizu, R., Harada, H., Minabe, H., Electronic control of car chassis present status and future perspective. *Convergence 88*, Toronto, Ontario, Canada, 1988, pp. 173–188.
60. Numazawa, A., Chassis systems electronics—Today and tomorrow. *ISATA*, Nr. 88005, pp. 77–100.
61. Aono, S., The next step in automotive electronic control. *Convergence 88*, Toronto, Ontario, Canada, 1988, pp. 83–89.

62. Schilkle, N., Fruechte, R., Boustany, N., Karmel, A., Repa, B., Rillings, J., Integrated vehicle control. *Convergence 88*, Toronto, Ontario, Canada, 1988, pp. 97–106.
63. Matsuo, Y., Harada, H., Yamamoto, M., Development of experimental vehicle with integrated chassis control. *JSAE Review*, 11(3), 30–36, 1990.
64. Wallentowitz, H., Scope for the integration of powertrain and chassis control systems: Traction control—All-wheel drive—Active suspension. *SAE 90*, Nr. 901168, Detroit, USA, 1990.
65. Dreyer, A., Heitzer, H., Control strategies for active chassis systems with respect to road friction. *SAE 91*, Nr. 910660, Detroit, USA, 1991.
66. Wallentowitz, H., Safety enhancing electronic control systems to aid drivers. *Automotive Technology International*, 92, 271–286, 1992.
67. Tanaka, H., Inoue, H., Iwata, H., Development of a vehicle integrated control system. *ImechE*, Nr. 925049, 1992, pp. 163–174.
68. Braess, H., Reister, D., The PROMETHEUS concept for future traffic—Conflict between dream and reality? *ImechE*, Nr. 925213, 1992, pp. 93–102.
69. Lugner, P., Mittermayr, P., Controlled additional 4-wheel steering at critical driving conditions. *AVEC'92*, Nr. 923045, Yokohama, Japan, 1992, pp. 245–251.
70. Ono, E., Hayashi, Y., Doi, S., Tanakami, K., Coordination of vehicle steering and suspension systems by integrated control strategy. *AVEC'92,* Nr. 923069, Yokohama, Japan, 1992, pp. 384–389.
71. Hirano, Y., Sato, Y., Ono, E., Tanakami, K., Integrated control system of 4WS and 4WD by H^{∞} control. *AVEC'92,* Nr. 923075, Yokohama, Japan, 1992, pp. 419–423.
72. Roppenecker, G., Wallentowitz, H., Integration of chassis and traction control systems, what is possible—What makes sense—What is under development. *Vehicle System Dynamics*, 22, 283–298, 1993.
73. Nagai, M., Ono, H., Model-following control for an integrated four-wheel steering and driving force control system taking into account tire nonlinearity. *JSAE Review*, 14(4), 35–41, 9307533, 1993.
74. Fujita, K., Ohashi, K., Inoue, Y., Ise, K., Development of automotive integrated control system. *EAEC6*, 1993, pp. 201–225.
75. Masinu, G., Babbel, E., Lugner, P., Margolis, D., Mittermayr, P., Richter, B., Integrated controls of lateral vehicle dynamics. *Vehicle System Dynamics*, 23, 358–377, 1994.
76. Lugner, P., Plöchl, M., Additional 4WS and driver interaction. *Vehicle System Dynamics*, 24, 639–658, 1995.
77. Shladover, S., Review of the state of development of advanced vehicle control systems (AVCS). *Vehicle System Dynamics*, 24, 551–595, 1995.
78. Beller, H., Rieth, P., Mit total chassis management auf dem Weg zum intelligenten Fahrwerk. XX. μ-*Symposium*, Bad Neuenahr, Germany, 10/2000, *Fortschr.-Ber. VDI-Reihe 12, Nr. 440*, Düsseldorf, North Rhine-Westphalia, Germany, 2000, pp. 48–76.
79. Trächtler, A., Integration der fahrdynamischen Funktionen durch Vehicle Dynamics Management (VDM). Fahrdynamikregelung, *Haus der Technik Conference Fahrwerktechnik*, Essen, Germany, September 20–21, 2001.
80. Trächtler, A., Liebemann, E., Vehicle dynamics management: Ein Konzept für den Systemverbund. *11. Aachener Kolloquium Fahrzeug- und Motorrentechnik*, Aachen, Germany, 2002.
81. van Zanten, A., Evolution of electronic control systems for improving the vehicle dynamic behavior. *AVEC'02, Proceedings 6th International Symposium on Advanced Vehicle Control*, Keynote Speech, Hiroshima, Japan, September 9–13, 2002.
82. Heinzl, P., Lugner, P., Plöchl, M., Stability control of a passenger car by combined additional steering and unilateral braking. *Vehicle System Dynamics Supplement*, 37, 221–233, 2002.
83. Schwarz, R., Muntu, M., Bauer, U., Schräbler, S., Tröster, S., Wein, M., Fritz, S., Maurischat, C., ESP II Fahrdynamik der nächsten Generation. Teil 1: Komponenten und Funktionen, *ATZ 11/2003*, *Automobil Technische Zeitschrift*, Jahrgang 105, pp. 1062–1069.
84. Schwarz, R., Muntu, M., Bauer, U., Schräbler, S., Tröster, S., Wein, M., Fritz, S., Maurischat, C., ESP II Fahrdynamik der nächsten Generation. Teil 2: Funktionsintegration und Elektronik, *ATZ 12/2003*, *Automobil Technische Zeitschrift*, Jahrgang 105, pp.1178–1182.
85. Köhn, P., Baumgarten, G., Richter, T., Schuster, M., Fleck, R., Die Aktivlenkung—Das neue fahrdynamische Lenksystem von BMW. *11. Aachener Kolloquium Fahrzeug und Motorentechnik*, Aachen, Germany, October 8–9, 2002.
86. Konik, D., Bartz, R., Bärnthol, F., Bruns, H., Wimmer, M., Dynamic Drive—Das neue aktive Wankstabilisierungssystem der BMW Group. *9. Aachener Kolloquium Fahrzeug- und Motorentechnik*, Aachen, Germany, 2000.

31 Active and Semi-Active Suspension Control

Davor Hrovat

CONTENTS

31.1 Introduction 1180
31.2 Performance Index 1182
31.2.1 Ride Comfort 1182
31.2.1.1 RMS Acceleration-Based Ride Metric 1182
31.2.1.2 Inclusion of Jerk 1182
31.2.1.3 ISO-Based Ride Metric 1183
31.2.1.4 NASA-Based Ride Metric 1183
31.2.1.5 Appropriate Choice of Ride Metric 1184
31.2.2 Rattlespace Design Constraint 1184
31.2.3 Tire Deflection Handling Constraint 1184
31.2.4 Combined Performance Index 1185
31.3 Passive versus Semi-Active 1185
31.4 External Factors: Road Roughness Description 1186
31.5 Optimal Suspensions for Quarter-Car Models 1189
31.5.1 One-DOF Models 1189
31.5.1.1 Some Useful Facts: Certainty Equivalence Principle 1191
31.5.1.2 Optimal Structure and Key Characteristics: Skyhook Damping 1192
31.5.1.3 Jerk-Optimal Case: Fast Load Leveling 1192
31.5.2 Two-DOF Models 1193
31.5.2.1 Illustrative Example 1196
31.5.2.2 Comparison with Conventional Passive Suspension 1196
31.5.2.3 Comparison with 1 DOF 1197
31.5.2.4 Influence of Unsprung Mass 1200
31.5.2.5 Additional Characteristics: Invariant Points 1200
31.5.2.6 Dynamic Vibration Absorbers 1200
31.5.2.7 Optimal Quarter-Car Structure 1202
31.5.2.8 Additional Considerations 1204
31.6 Optimal Suspensions for Half-Car Models 1204
31.7 Optimal Suspensions for Full-Car Models 1208
31.8 Related Topics 1211
31.8.1 Semi-Active Suspensions 1211
31.8.1.1 On–Off Semi-Active Damping 1211
31.8.1.2 Continuously Variable Semi-active Damping 1211
31.8.1.3 Semi-Active Spring 1212
31.8.2 State Estimation and Closed-Loop System Robustness 1213

31.8.3 Nonlinear Suspension Systems and Controls 1213
31.8.3.1 Nonlinear Control Design through Optimization 1214
31.8.3.2 Nonlinear Controls as an Extension of LQ Controls 1214
31.8.3.3 Other Approaches to Nonlinear Suspension Control 1214
31.8.4 Practical Considerations 1216
31.8.5 Other Aspects 1217
Acknowledgments 1218
31.A Appendix: Nonlinear Suboptimal Control 1218
References 1220

31.1 INTRODUCTION

Vehicle suspensions perform a number of important ride, handling, and overall vehicle agility and safety-related functions. They must support the vehicle sprung mass and secure the proper posture for both the vehicle and the driver. At the same time, they must provide effective isolation from the road roughness-induced disturbances.

It is desirable to maintain a proper vehicle posture in the presence of various inertial and external forces and moments caused by braking, turning, wind gust, and other events. For example, during braking and rapid accelerations in turns, a proper anti dive and anti squat/anti pitch suspension geometry and anti roll bar design is typically used to provide a good vehicle posture and control for most of current vehicles equipped with conventional, passive suspensions. This in turn can facilitate more comfortable and more "in-control" driver posture, leading to better handling and overall safer vehicle ride. This can be further enhanced with active suspensions discussed later.

There are many factors influencing ride comfort. These include road roughness, curvature, and undulation; seat configuration; vehicle interior, color, and spaciousness; noise, vibration, and harshness (NVH); or even odor and visual perception of surrounding scenery. The most relevant for ride studies is ground input disturbances caused by prevailing road roughness, which will also be the main focus of the present survey. Because of this, in a certain sense the ride design problem could be viewed as an equivalent filter design problem where one would attempt to eliminate the negative effects of vibrations caused by road roughness. At the same time it would be desirable to pass the low-frequency components of road inputs that are necessary to follow the terrain, while satisfying the given design and system-imposed constraints.

Conventional vehicle suspensions achieve the road-induced vibration isolation through passive means such as springs and dampers or shock absorbers. On the other hand, the active suspensions are characterized by a requirement that at least a portion of suspension force generation is provided through active power sources such as compressors and hydraulic pumps. Practical applications of active suspensions have been facilitated by maturing of microprocessors and associated electronics, which in turn also influenced actuator and sensor developments.

Some of the potential benefits of active suspensions were predicted decades ago by the first pioneer researchers. Indeed, the optimal control techniques that were "launched" with Sputnik and used in the aerospace industry since the 1950s and 1960s have also been applied to the study of active suspensions, starting from about the same period [1]. Early investigations [2–5] used the simplest possible, one-degree-of-freedom (1-DOF), quarter-car models. The resulting optimal suspension design problem fits naturally into the linear–quadratic–Gaussian (LQG) or, equivalently, the Wiener–Hopf setting [4]. It will be argued later that, despite their simplicity, these early contributions had a profound impact on the practical implementation some 15–20 years later, when the utilization of these theoretical results was made possible by the ever-increasing capacity of microcomputers in the automotive industry.

It should be pointed out that the design and synthesis of active suspensions can be approached from many avenues: modal analysis [6]; classical techniques, such as root locus and Bode and Nichols plots [7–9]; eigenvalue assignment [10]; Youla parameterization and linear fractional

transformation [11]; simulation using multibody dynamics computer models [12–15]; design of experiments [16]; bond graph modeling methodologies [17]; optimization—e.g., nonlinear programming and multicriteria optimization [18–20] and optimal control [2–5,21–24]; and, more recently, neural networks, fuzzy logic, evolution strategy, and learning automation [25–30]. While each of these approaches can bring some useful perspectives and insights, this chapter will focus on the applications of optimal control techniques and especially on the related work by the author and his colleagues at Ford. It will be demonstrated later that this constitutes a very natural approach revealing important global trends and insights such as the optimal system structure, performance potentials and trade-offs, and fundamental actuator and sensor requirements.

Specifically, it will be shown that with this approach even the simplest possible, 1-DOF linear vehicle models can produce very useful insights. For example, a straightforward application of the well-known linear–quadratic (LQ)-optimal control methodology that minimizes vehicle root-mean-square (rms) acceleration as a typical comfort measure will lead to the concept of "skyhook" damper introduced in Section 31.5.1. The latter has found many practical applications in the automotive industry, especially in the context of semi-active (SA) damping. Further introduction of an rms jerk term as an additional comfort measure leads to the concept of fast load leveling that offers additional practical advantages for active suspensions.

Introduction of one additional DOF representing the wheel-hop dynamics leads to the 2-DOF quarter-car model that was thoroughly investigated by a number of researchers resulting in a new insight, especially during the 1980s. For example, as it will be shown later in Section 31.5.2, the addition of the wheel-hop mode results in one more vehicle handling-related constraint and additional invariant points in related transfer functions (e.g. the transfer function between ground input displacement and sprung mass acceleration). In turn, this significantly reduces the potential benefits of active suspensions. Once the nature of this constraint is well understood, one way to recover a good portion of this lost performance is via structural changes using appropriate dynamic vibration absorbers.

Next, moving from 1D to 2D models, it will be possible to use the front wheels to preview the road disturbance that will eventually impact the rear wheels. This naturally leads to the concept of preview controls, which, as it will be shown in Section 31.6, can offer further significant benefits. Finally, moving to the full 3D models, it will be shown that under reasonable assumptions, many of these models can be further decomposed to simpler 1D and 2D counterparts.

It should be pointed out that for the earlier system-level, global study of potential benefits, it is quite appropriate to use simple, linear models, where the actuator dynamics are typically neglected. Again, the main objective is to establish what the best possible performance improvement is, that active suspension can supply under ideal conditions. Obviously, if these maximum possible potential improvements are not significant, then there is little incentive to further pursue a particular active setting all the way toward practical implementation. At the same time, these types of studies naturally lead to consideration of structural changes needed to achieve certain performance, that is, they facilitate outside-the-box thinking. Once the potential benefits have been established for a given active suspension configuration, it is then appropriate to start considering different actuation schemes that could deliver these projected benefits. At this stage, a careful consideration of actuator dynamics and design is necessary, since the actual process of delivering the projected benefits is by no means trivial in practice.

The earlier linear setting is quite appropriate for typical ride problems characterizing on-road vehicle operating at a given speed on typical roads with a prevailing roughness that excites relatively small motions of vehicle masses and suspension components. This situation changes for off-road operations and/or under occasional large bump or pothole excitations. In this case, nonlinear dynamic extensions need to be considered too. One such extension that naturally builds on the earlier linear studies will be given in Section 31.8.3.2. This will be followed by consideration of important practical issues and related topics.

The next three sections summarize the main ingredients needed to formulate the optimal suspension design problem to be considered in the rest of the chapter. The necessary infrastructure includes appropriate definitions for ride performance metric, active and passive suspensions, as well as the corresponding representation for road roughness, which acts as the main disturbance. The high-level active suspension design and investigation of performance potentials and trade-offs at the system level is addressed in the subsequent section, which also contains a brief description of related vehicle models.

31.2 PERFORMANCE INDEX

The ground vehicle suspension design is influenced by a number of often conflicting factors and attributes. Good vibration isolation is required to secure the occupants' ride comfort, whereas good posture and road holding is important for vehicle handling, which in general leads to enhanced safety and overall vehicle control and stability. Key design constraints are represented by maximum allowable relative displacements between the vehicle body and various unsprung mass components, including wheels, bump stops, and protruding parts of the steering mechanism. Additional constraints are imposed by the overall system robustness, reliability, weight, and cost requirements. Many of these factors will be discussed next and then combined in an overall performance index (PI) that will be used for most optimal control designs in later sections.

31.2.1 Ride Comfort

Due to its highly subjective nature, there are many different ways to quantify ride comfort (see Chapter 36), which is typically measured at the suitable sprung mass (or vehicle main mass) location such as the driver's seat or vehicle body center of mass (CM) or CG location. They vary in complexity from the simple rms and maximum vertical or heave acceleration metrics to frequency-weighted metrics and much more complex measures that combine contributions from different vehicle vibration modes. In the next few subsections, we will summarize the most relevant metrics and associated characteristics. Additional details and approaches can be found elsewhere in this handbook (see Chapter 36).

31.2.1.1 RMS Acceleration-Based Ride Metric

One of the simplest ride measures is based on calculations of rms values of vertical acceleration typically measured at the driver's or passengers' seat locations. In a related 1978 Ref. [31], the authors conducted a field study involving two different automobiles, 78 passengers, and 18 different road sections to conclude that "excellent correlation was found to exist between the subjective ride ratings and simple rms acceleration measurements at either the vehicle floorboard or the passenger/seat interface." The main result from this work is reproduced in Figure 31.1, which shows the correlation between the average or mean personal ratings and rms acceleration measured for combined vertical and lateral directions, where higher ratings imply better ride comfort. A similar correlation was obtained for vertical-only accelerations.

31.2.1.2 Inclusion of Jerk

Further refinements of the rms ride measure are possible through introduction of mean-square (ms) jerk, which is the derivative of the acceleration. Some authors [32–34] have advocated the inclusion of jerk as an additional comfort measure that amplifies the contribution from high-frequency disturbances, which are typically lumped under NVH. In Section 31.5.1.3, we will make use of this additional term to arrive at a very desirable basic optimal suspension structure.

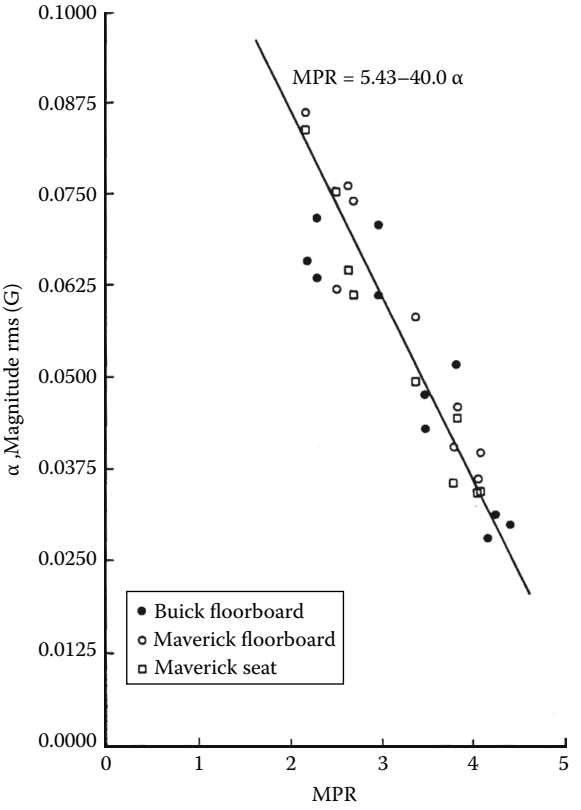


FIGURE 31.1 Experimental rms accelerations versus mean personal ratings and corresponding least squares curve fit [31].

31.2.1.3 ISO-Based Ride Metric

The rms (or ms) quantifiers do not reflect the frequency dependence of human sensitivity to vibrations. To account for this aspect of ride comfort, the International Organization for Standardization (ISO) has developed a standard ISO 2631, which also takes into account the length of time of human exposure to the vibrations [35]. This is represented through the plot of ride comfort boundaries as a function of third octave band center frequency. According to the earlier standard, the region of greatest human sensitivity to vertical vibration lies between 4 and 8 Hz, which roughly includes various resonances of human internal organs (see Section 36.2.2). Although developed primarily under the assumption of sinusoidal or periodic excitation, the standard also applies to the case of broadband vibrations, which are represented via rms values within prespecified one-third octave bands [36,37]. Similar considerations apply for vibration in the lateral direction with respect to the human body. Although not explicitly used here, the ISO-based metric could be easily incorporated in the proposed optimal control framework via an appropriate state space augmentation.

31.2.1.4 NASA-Based Ride Metric

The earlier measures are primarily based on decoupled tests, typically performed in only one direction at a time. A more comprehensive measure, which takes into account interdependency of various modes of vibrations, was developed by NASA, primarily for aerospace applications [38]. A similar measure was adapted and extended to automotive applications as summarized in [39], where the interdependence between heave, pitch, roll, and many other factors have been considered. This metric can be used for optimal suspension designs where vehicle models have many DOFs (in addition to the standard heave mode), as well as for more detailed vehicle evaluations of different suspension concepts and strategies.

31.2.1.5 Appropriate Choice of Ride Metric

Which of the earlier ride metric alternatives one should evoke will depend on the context of their usage. If the primary use is in the context of simple models, then the simple rms-based metric will suffice. On the other hand, if models with higher DOF are used, then the more complex measures such as the earlier NASA-based metric may be appropriate.

To date, there is a lack of comprehensive comparison among different ride comfort measures. An earlier comparison [31] between the simple ms measures and the more involved ISO standard 2631 was inconclusive; indeed, the correlations with ride ratings even resulted in a slight advantage for the simple rms metric. In view of this and in the interest of the simplicity of approach that is often needed to reveal the main trends, it is natural to start with the ms acceleration as a measure of ride comfort. This also matches the relatively simple models of relevant vehicle dynamics and the road input disturbances that will be discussed later. Once sufficient information has been obtained about the potential benefits at this simplified or idealized system level, it is appropriate to consider other, more detailed metrics—such as 2.1.3 and 2.1.4—along with corresponding increase in model and optimization process complexity and level of detail.

31.2.2 Rattlespace Design Constraint

Based on the previous section, if all that we wanted was to minimize the ms vertical acceleration of a sprung mass, then the solution would be conceptually trivial: just produce a constant suspension force equal to the sprung weight, and the acceleration in question will always be zero (note that while conceptually trivial, it is by no means trivial to actually implement even this simple strategy, as will be discussed elsewhere in this handbook—see Chapter 22). The problem with such a "solution" is that the affected vehicle would then need unrealistically large suspension strokes to negotiate large road irregularities and hills and valleys, that is, in effect the vehicle would have to transform in to a flying object! The fact that only a limited suspension stroke is available must be reflected by incorporating at least one additional constraint into the ride optimization problem.

This additional, design-related constraint can be represented as either a hard or soft constraint in the sense that is used in optimization theory. That is, the "soft" constraint is typically introduced via an integral penalty term thus allowing for occasional violation of the constraint. The "hard" constraint, on the other hand, is strictly enforced, typically via an explicit (in)equality. The hard constraint approach is more accurate, although even in this case there is no guarantee that comprehensive, accurate, optimal solutions can be found. Simply, it is practically impossible to predict all road irregularities that will be encountered throughout the life of a large production volume of vehicles. Because of this, and because of analytical convenience, it is more common to introduce a soft constraint in the form of ms suspension stroke or so-called rattlespace. The resulting PI to be optimized, then, typically consists of two terms:

- The ms acceleration, which is a measure of ride comfort
- The weighted ms rattlespace, which reflects design and packaging constraints

This simple, two-term PI will be used in the context of the simplest possible, 1-DOF quarter-car vehicle models, to be discussed in Section 31.5.1. A similar two-term PI can be and was also used for the 2-DOF, quarter-car models. However, the three-term PI incorporating vehicle handling constraint (to be introduced next) is more suitable in this case, since for many on-road vehicle operations, the handling constraint is more limiting (or "active") than its rattlespace counterpart.

31.2.3 Tire Deflection Handling Constraint

For slightly more complex, 2-DOF quarter-car models, it is possible to impose one additional, handling-related constraint. This pertains to tire wheel-hop dynamics, typically manifested as a relatively lightly damped mode at 10–12 Hz. This can result in considerable time variations of the

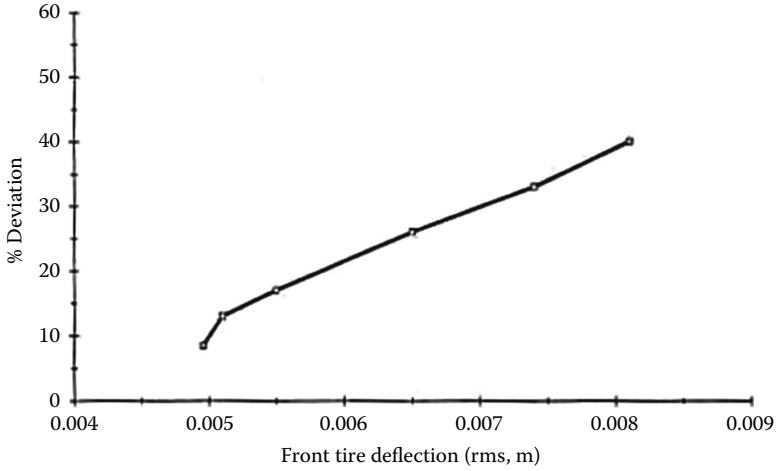


FIGURE 31.2 Percent change in path deviation versus change in tire deflection for the case of traversing a rough road with crosswind gust.

tire normal force around (quasi) static equilibrium, which in turn can lead to a loss of contact with the ground and reduced handling ability. The detrimental effect of tire force variation on vehicle handling comes in good part from the convex relationship between tire cornering forces (lateral forces) and normal forces, for a given tire slip angle (see Chapter 17). To investigate these effects, the following study was conducted [40]: A vehicle traveling at 100 km/h with a constant straight-ahead steering was subjected to a crosswind gust of 36 km/h on a rough road. Figure 31.2 shows the resulting lateral deviation in vehicle trajectory after 4 s that is plotted as a function of tire rms deflection around equilibrium. It can be seen that there is an almost linear relationship between the vehicle path deviation—an explicit handling metric—and the tire rms deflection, an implicit metric. Thus the latter can be used next to include the handling in a three-term PI representing vehicle ride, handling, and design/packaging requirements.

31.2.4 Combined Performance Index

In view of the preceding text, we can now formulate the final three-term PI as follows: (1) For ride comfort use the ms acceleration term; (2) for rattlespace constraint use the weighted ms rattlespace, which reflects design and packaging limitations; (3) and for handling limitations, use the weighted ms tire deflection. This three-term PI will be used in the context of 2-DOF quarter-car vehicle models, to be discussed in Section 31.5.2. As a final word of caution, it should be pointed out that some authors have attempted to use only a two-term PI consisting of the earlier ms ride and handling metrics. This leads to pathological situations resulting in infinite rattlespace requirements.

31.3 PASSIVE VERSUS SEMI-ACTIVE

Different suspensions satisfy the earlier mentioned ride, handling, and design requirements to differing degrees. Although significant improvements can result from a designer's ingenuity, on the average, suspension performance is mainly influenced by the type or class of suspension used. Here one distinguishes, in ascending order of improved performance, between passive, SA, and fully active suspensions. Passive suspensions are those found on most conventional vehicles. Roughly, they are characterized by the absence of external power sources, whereas the active suspensions require additional sources, such as compressors and pumps, to achieve superior ride and handling performance. The SA (or semi passive) suspensions [41] then fall in between passive and active, since they essentially act as time-varying dampers and thus offer potentially significant performance improvements while requiring relatively small or negligible external power supplements.

More precise definitions of active and passive suspensions were given in [42,43] following similar definitions of passivity and passive operators as used, for example, in electrical networks and mathematics. According to the latter, an operator P is passive if there exists some constant k such that the inner product

$$< Px \mid x >_T > k \tag{31.1}$$

where the inequality must hold for all final times T and all elements $x(\cdot)$ of an extended inner product space *He,* composed of all functions that do not have a finite escape time (i.e., do not "explode" in a finite time). If one associates x with the actuator velocity and Px with the corresponding force, then inequality (31.1) reflects a power dissipation requirement for passive actuators.

For the case of linear time-invariant dynamic systems, one can assume, without loss of generality, that the constant k equals 0. Moreover, in this case the passivity inequality (31.1) is equivalent to requiring that the associated transfer function matrix $P(s)$ be a Rational Positive Real (RPR) matrix. The RPR matrix $P(s)$ must satisfy the following three requirements [42,44]:

1. All elements of $P(\cdot)$ have poles in the left-hand side plane, that is, where $\mathrm{Re}[s] <= 0$.
2. The Hermitian matrix $P(j\omega) + P^*(j\omega)$ is positive semi-definite for all $\omega \in \Re$, with jω not a pole of any element of $P(\cdot)$.
3. If $j\omega_o$ is a pole of any element of $P(\cdot)$, it is at most a simple pole. In that case the residue matrix,

$$P_o = \lim_{s \to j\omega_o} (s - j\omega_o) P(s) \text{ in case } \omega_o \text{ is finite and}$$

$$P_\infty = \lim_{\omega \to \infty} P(j\omega)/j\omega \text{ in case } \omega_o \text{ is infinite, is positive semi definite Hermitian}$$

where P^* stands for the complex-conjugate transpose and the elements of P belong to the class of distributions or generalized functions. Note that the immittance matrix of a linear time-invariant, lumped, finite, and passive n-port is necessarily RPR. In this case, an alternative description of passivity can be obtained via the state-space approach, which leads to a Positive Real Lemma or, as is often called, "Kalman–Yakubovich Lemma." The resulting set of matrix equations is then equivalent to (1–3) in the earlier mentioned requirements and is sometimes preferred in the synthesis of passive networks.

The earlier definition was used to check the passivity of the relationship between optimal suspension forces and velocities for simple 1-DOF [43] and 2-DOF [45] vehicle models. Similar approach can be used to explicitly constraint the suspension synthesis to the subclass of passive structures or to constraint the amount of active energy spent by active suspensions with regenerative capabilities during given road-induced maneuvers. This offers some interesting and challenging future avenues of research.

31.4 EXTERNAL FACTORS: ROAD ROUGHNESS DESCRIPTION

There are two main sets of disturbances that affect vehicle ride and handling: one is caused by road roughness and the other by different forces and moments that originate in various inertial and aerodynamic loadings, as caused by braking, turning, and wind gusts, for example. The most relevant for ride studies are ground input disturbances caused by road roughness.

The road input disturbances can be described in many possible ways. Chiefly, they can be classified as shock or vibration. Shocks are discrete events of relatively short duration and high intensity, as, for example, caused by a pronounced bump or pothole on an otherwise smooth road. Vibrations, on the other hand, are characterized by prolonged and consistent excitations that are felt on "rough" roads. A well-designed suspension must perform adequately in a wide range of shock and vibration environments.

In the context of vibrations, the road roughness is typically specified as a random process of a given displacement power spectral density (p.s.d.). Typical measured spectral densities of various terrains from [46] are reproduced in Figure 31.3. An often used approximation of these road displacement p.s.d. curves is given in the form

$$S(\Omega) = A\Omega^n \tag{31.2}$$

where

Ω is the spatial frequency, typically in units of "radians per length"

A and n are appropriate constants

For the most commonly used case of $n = -2$, the corresponding line with the slope of -2 in log–log scale is shown in Figure 31.3 as well. With $n = -2$, the displacement spectra of Equation 31.2 imply a white-noise ground input *velocity,* which—when augmented by commonly evoked Gaussian assumption—conveniently matches the well-known, standard LQG assumptions for process noise. A procedure to construct and preprocess such a road sequence on a digital computer is given in [47].

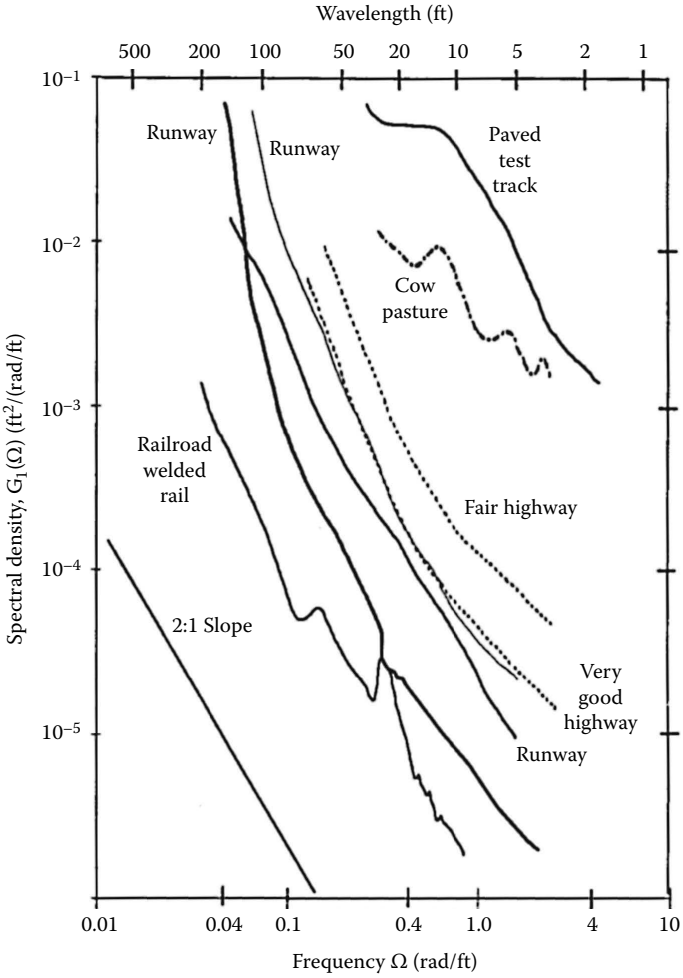


FIGURE 31.3 Power spectral density of various roads and terrains.

The simple expression (31.2) approximates various roads with different degrees of fidelity [39]. An example, the so-called Rochester Road section of low-speed track at the Ford Dearborn Proving Ground, is shown in Figure 31.4a, reproduced from [39]. Here, it can be seen that the standard white-noise velocity assumption, represented by the thick line, fits the data very well ($n = -2.02 \sim -2$), except possibly at the very low frequencies.

On the other hand, Figure 31.4b shows a different road section from [39]. This is the so-called section between broken road and Rochester Road that displays significant deviations, especially at low frequencies, from the fitted model of Equation 31.2, where $n = -1.99$. A better correlation with measured spectra in this case was obtained with the help of more complex, exponential fit [24]. Other variants of road description include various forms of processing of "raw" white-noise velocity spectra. This processing ranges from a simple, first-order filtering to third-order filtering used to capture possible strong oscillatory components in the spectra [39].

The extensions of these single-track models toward spatial road surfaces are described in [39] and references therein. The authors propose to consider a road model illustrated by Figure 31.5, where left and right track inputs are derived based on the road roughness at suitable points in between the

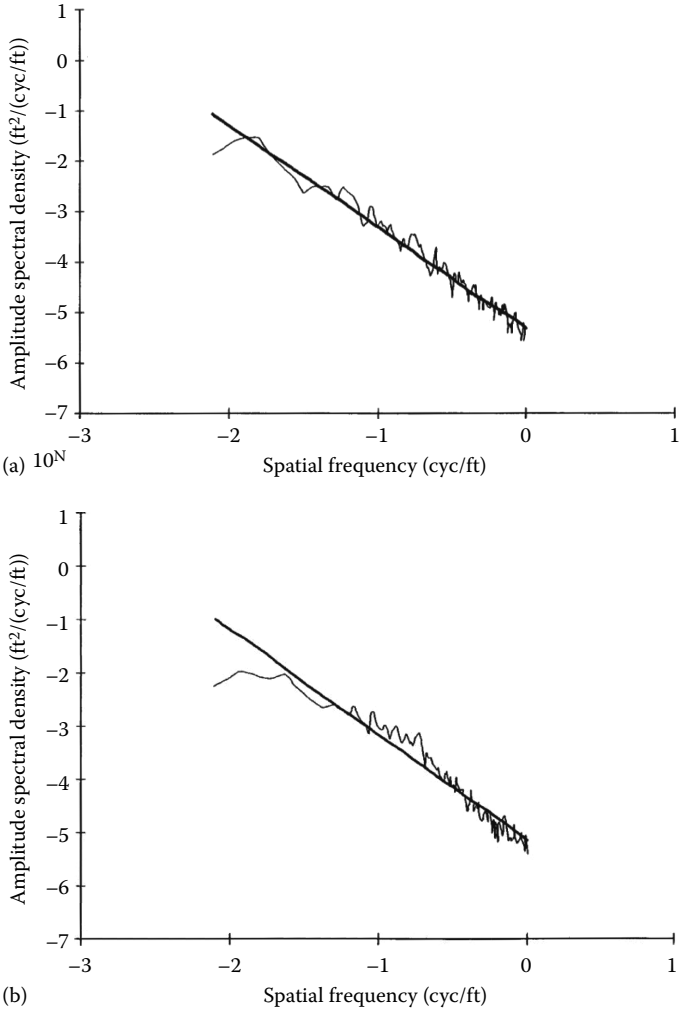


FIGURE 31.4 (a) Comparison of model given by Equation 31.2 with $n = -2.02$ and measured road roughness for "Rochester Road section." (b) Comparison of model given by Equation 31.2 with $n = -1.99$ and measured road roughness for "section between broken road and Rochester Road."

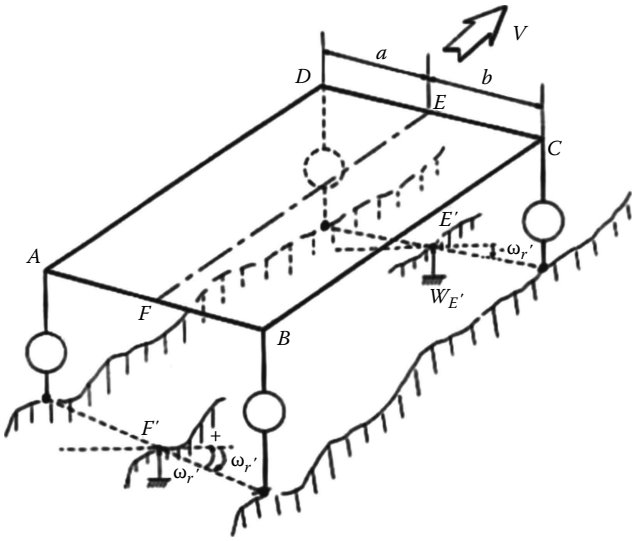


FIGURE 31.5 Road configuration for 3D vehicle model studies.

tracks, usually along the vehicle centerline represented by E' and F' in Figure 31.5. Also given in Figure 31.5 are the relevant expressions for wheel–road contact velocities as a function of vertical (W_E and W_F) and rotational (ω_E and ω_F) velocities at E and F. It is furthermore assumed that W_E and ω_E are (low-pass filtered) white-noise velocity processes and that W_F and ω_F are their delayed versions. The resulting, cross-correlated road input representation is suitable for 3D optimal control design and analysis and, as will be shown later, leads naturally to a particular form of decoupling and simplification of the corresponding full-car problem.

31.5 OPTIMAL SUSPENSIONS FOR QUARTER-CAR MODELS

This section reviews some of the many findings based on simple, quarter-car or "unicycle," 1D vehicle models. These have been the subject of many investigations during the past few decades. The most appropriate approach here is to start with the simplest possible, 1-DOF models, which, while neglecting the unsprung masses, still yield some very useful results, forming the bounds for limiting optimal performance of more complex, higher DOF models.

31.5.1 One-DOF Models

In view of Section 31.2, the optimal suspension design problem for the 1-DOF model of Figure 31.6 can be expressed as follows:

$$\text{Minimize}\left\{PI_1 = E\left(x_1^2 + ru^2\right)\right\} \tag{31.3}$$

subject to the state equations

$$\dot{x}_1 = x_2 - w \tag{31.4}$$

$$\dot{x}_2 = u \tag{31.5}$$

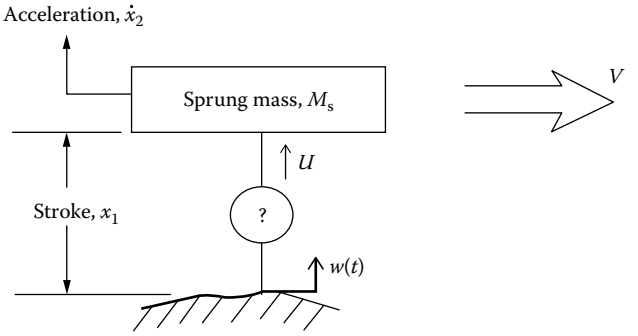


FIGURE 31.6 A 1-DOF model.

where the earlier expectation represents steady-state ms values. The states x_1 and x_2 from Figure 31.6 represent rattlespace (suspension deflection with respect to equilibrium) and vehicle vertical velocity with respect to an "inertial" ground—in practice, a low-frequency "smooth" component of road inputs. We assume that both states are measurable and available for control; other cases are discussed in 31.8.2. The rattlespace constraint on x_1 is introduced in the PI (31.3) to reflect a limited suspension stroke capability. As discussed in Section 31.4, the ground input velocity w due to road roughness effects is typically modeled as a white-noise, Gaussian process specified via

$$E[w(t)] = 0 \tag{31.6}$$

$$E[w(t_1)\,w(t_2)] = 2\pi W\delta(t_1 - t_2) \tag{31.7}$$

where $\delta(\cdot)$ stands for the impulse or Dirac delta function and W is the two-sided road velocity power spectral density (see [47]),

$$W = AV$$

where

A is the road roughness factor
V is the vehicle forward velocity

The optimal suspension design problem then consists of minimizing the PI (31.3), which penalizes excessive suspension stroke and sprung mass acceleration u, subject to the state Equations 31.4 and 31.5. The acceleration is a convenient control variable in the present setting, and it amounts to normalizing the original control force U from Figure 31.6 with respect to the sprung mass m. The weighting parameter r in Equation 31.3 acts as a "tuning knob," so that larger r results in smaller accelerations, that is, improved ride comfort, and increased suspension stroke requirements.

For the present second-order system, the LQ-optimal solution can be calculated analytically ([24]) from the associated Riccati equation to yield the optimal control acceleration:

$$u_{LQ} = -r^{-1/2}x_1 - \sqrt{2}r^{-1/4}x_2 \tag{31.8}$$

The corresponding optimal trade-off curve is shown as a solid line in Figure 31.7. Both the rms acceleration and rattlespace have been conveniently normalized with respect to road/speed characteristics, so that $\tilde{x}_1 = x_{1,rms}/\sqrt{2\pi AV}$. Note in particular that traveling on rough roads, and/or at high speeds, implies *smaller* normalized rattlespace and larger normalized accelerations. The solid

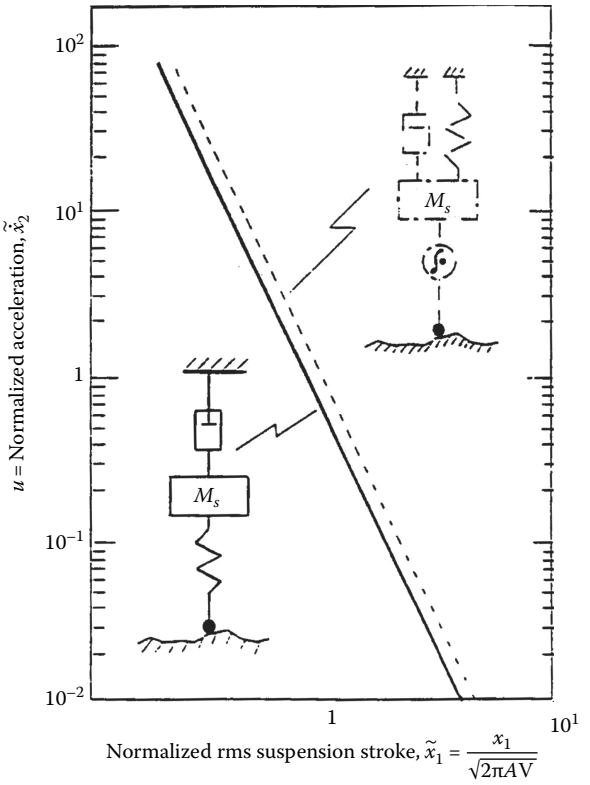


FIGURE 31.7 Optimal performance trade-offs for 1-DOF model with PI_1 (solid line) and PI_2 (dashed line).

straight line in Figure 31.7 is parameterized by the weighting parameter r and has a slope of −3 on the log–log scale, which implies that each 10% increase in allowed rattlespace results in ~30% decrease in rms acceleration. Larger values of r result in softer suspensions, with smaller accelerations and larger rattlespace requirements.

31.5.1.1 Some Useful Facts: Certainty Equivalence Principle

Note that the assumptions in Section 31.4 imply that all response variables are Gaussian. This makes it possible to calculate the probability that the suspension stroke and other variables of interest will remain within certain bounds around the equilibrium. For example, evoking the standard results for Gaussian or normal distributions, it can be concluded that the variation of these variables around their equilibria will exceed three times their corresponding rms values during only 0.3% of ride time. It should also be remarked that, because of the central limit theorem [24], similar conclusions apply for non-Gaussian road sequences. This was also observed by a number of researchers (e.g., [39,48]) based on measured road data.

While the earlier problem could be also addressed via optimal filtering [24], one of the advantages of the LQ approach is that it also applies to linear, time-varying systems, and there are many well-developed numerical techniques for its solution in the most general multi-input–multi-output (MIMO) setting. Moreover, one can evoke the certainty equivalence principle (CEP) [24], which states that the optimal controller for the original stochastic problem coincides with the solution of an equivalent deterministic problem, obtained by replacing all random variables by their expected values. In the present case, it can be shown [49] that the deterministic problem amounts to replacing the white-noise ground velocity by an equivalent step in ground displacement, that is, by an impulse in ground velocity. Alternately, the ground step can be replaced by an equivalent nonzero state (rattlespace in the case of 1-DOF model) initial condition, which occurs when dealing with optimal shock isolation during,

for example, an airplane landing. The main point is that with the "one shot" LQ approach, one can solve two problems: one is the stochastic, vibration isolation problem with the PI given as in Equation 31.3, and the other is the corresponding deterministic or shock isolation problem [49].

31.5.1.2 Optimal Structure and Key Characteristics: Skyhook Damping

The optimal suspension structure can be deduced directly from the LQ problem formulation knowing that the LQ-optimal control is a linear function of states. It consists of a linear spring placed between the vehicle body and the road and a so-called skyhook damper [2,41] placed between the vehicle sprung mass and an inertial, smooth ground. This is illustrated by a schematic shown on the left side of Figure 31.7. Note that the optimal structure can be predetermined even before pursuing any actual (numerical) solution of the problem. The same can be said about equivalent nonlinear problems, where the skyhook damper would now be characterized by a nonlinear force description. Since an absolutely smooth inertial ground is in practice not available from a moving vehicle, the actual suspension will have to be placed between the vehicle and the road, which in effect acts as a "moving ground." It can then be shown, using the previously discussed definition of passive systems, that the LQ-optimal suspension for the stochastic case is necessarily an active device, for any value of the weighting parameter r [43].

In addition, it follows from Equation 31.8 that the optimal closed-loop system is mildly oscillatory with a favorable damping ratio of ~0.7. To put this in proper perspective, it should be pointed out that the corresponding ratios of typical conventional passive suspensions are around 0.2–0.3. As will be shown in Section 31.5.2.2, this range of values is close to optimal for the passive case: smaller values would amplify the body resonance, whereas larger values would increase disturbance force transmissibility from road through the damper. This is where the skyhook dampers fundamentally differ from their passive counterparts: by emulating a contact with the hypothetical, smooth, inertial ground, the optimal algorithm facilitates much larger damping rates leading to a well-damped body mode with no negative transmissibility-related side effects. The net effect is a dramatic improvement in ride and handling for road/steering excitation frequencies around the body modes. For example, the dynamic sag due to body roll in turns can be two to three times smaller with the optimal active suspension of the same stiffness as its passive counterpart [24]. Thus, even though the handling per se was not taken into account in the earlier optimization, the resulting optimal structure nevertheless offers significant handling benefits. While the present observations are based on the simple 1-DOF model, most of the facts also apply to the full 3D setting. For example, the optimal damping ratio of 0.7 applies for the key rigid body modes (heave, pitch, and roll) characterizing the dominant, sprung mass mode of oscillations (see Section 31.7).

31.5.1.3 Jerk-Optimal Case: Fast Load Leveling

Since there is evidence that the jerk, as well as acceleration, is important for ride comfort (see Section 31.2.1), it is appropriate to consider extending the earlier 1-DOF model by including the sprung mass jerk, $\dot{u}$ (the derivative of acceleration u) as an additional comfort measure in the PI of Equation 31.2. Following [49,50], we define

$$PI_2 = E\left(x_1^2 + r_1 u^2 + r_2 \dot{u}^2\right) \tag{31.9}$$

The problem can next be transformed into a more standard, LQR form by defining the sprung mass acceleration u as an additional third state and the sprung mass jerk, $\dot{u}$, as the new control input. There are now two tuning parameters, r_1 and r_2, which weigh ms acceleration and jerk, respectively.

As shown in the earlier references, it is still possible to obtain the analytical solution for the resulting third-order LQR problem. The optimal structure in this case is shown with dashed lines in Figure 31.7. In addition to a skyhook damper, this structure includes a skyhook spring and a load leveling device that operates on the integral of suspension deflection. This configuration in fact has some desirable characteristics.

The load leveling secures zero steady offset in the case of external load disturbances caused by vehicle weight changes and inertial forces in turns or during braking, for example. Being fully integrated in the LQR synthesis process, this load leveling action can be relatively fast, with settling times of the order of 1–3 s [49,50], as opposed to more conventional slow-acting load levelers requiring dozens of seconds to settle. While the fast load leveling facilitates load disturbance rejection and good tracking of low-frequency components of the road, the skyhook damper and spring minimize transmissibility of road-induced disturbances.

The corresponding optimal ride performance is represented by the dashed line in Figure 31.7, which applies for the extreme case when $r_1 = 0$. This results in minimal possible jerk for a given rattlespace. Even in this extreme case, there is a relatively modest, about 10%, increase in rattlespace requirements for the same level of rms accelerations [50]. At the same time, while the optimal damping ratio of 0.7 is now reduced to 0.5, it is still substantially above the typical passive values of 0.2–0.3.

On the other hand, at the other extreme where $r_2 = 0$, for the "standard" 1-DOF case, the rms jerk is theoretically infinite or—taking into account the wheel-hop and other neglected dynamics—very large. This implies that substantial reduction in jerk will be obtained with the new jerk-optimal structure. In practice, this means that the high-frequency, NVH isolation of the proposed jerk-optimal suspension will be much superior to the one achievable with the standard structure.

The basis for this improved low-pass filtering is in the fact that the conventional spring in the standard configuration has been replaced by an integrator and a skyhook spring, both reducing the transmission at high frequencies. The improved high-frequency filtering in this case also leads to less stringent desired force bandwidth requirements. This is a favorable consideration for narrow bandwidth or "slow-acting" active suspension actuators, which offer a promising approach to practical production implementation of active suspensions.

In summary, the application of optimal control technique to a single, 1-DOF, quarter-car or "unicycle" model has resulted in fundamental insights about the optimal suspension structure and performance potentials. This includes the concept of a skyhook damper, which has found applications in most of the SA and active suspensions in production or under research and development. Although the simple model neglects the wheel-hop mode, the earlier principal isolation structure is still near optimal and, as needed, can be augmented by a damper in parallel with the active actuator. Typically, this configuration is often used in practice to control the 2-DOF case. It can be shown [24] that this strategy is usually not far from optimal. Inclusion of jerk as an additional ride metric leads to the concept of fast load leveling and associated integral control, which can also be extended to more complex model settings. To this end one could also explicitly include the integral term(s) by appropriately augmenting the state-space vehicle model formulation [51].

31.5.2 Two-DOF Models

A typical 2-DOF model structure including wheel-hop dynamics is shown in Figure 31.8a, along with its 1-DOF active (Figure 31.8b) and 2-DOF passive (Figure 31.8c) counterparts. The related model equations are given in Chapter 22 on related advanced suspension hardware. This allows for inclusion of a handling measure in the PI in the form

$$PI_3 = E\left(r_1 x_1^2 + r_2 x_3^2 + u^2\right) \tag{31.10}$$

where x_1 and x_3 are the states representing tire and suspension deflection as per Figure 31.8. As elaborated in Section 31.2.3, the tire deflection is an indication of road-holding ability, in the sense that smaller tire deflections lead to better vehicle handling characteristics. Since this term in Equation 31.10 represents an additional "soft" constraint, it is to be expected that this will lead to deterioration in performance as measured by rms acceleration and rattlespace.

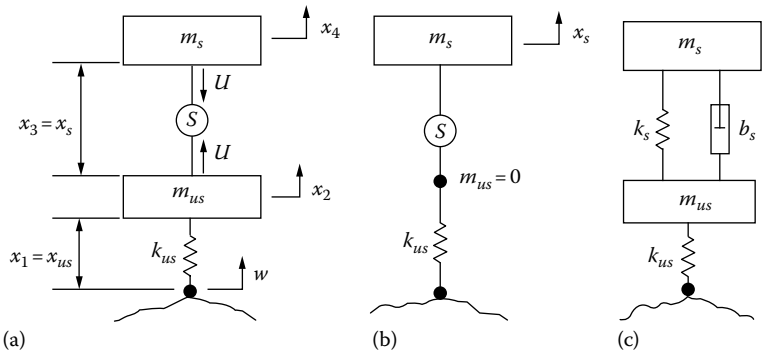


FIGURE 31.8 (a) A 2-DOF vehicle model with active suspension *S*, (b) corresponding limiting case when $m_{us} = 0$, (c) 2-DOF model with conventional passive suspension.

The earlier four-state LQ problem was solved numerically [24], and the results of the corresponding global study for various r_1 and r_2 combinations are shown in Figures 31.9a and b [57] in the form of "carpet" plots of normalized rms acceleration versus suspension and tire deflections, respectively. Here, the ratio of sprung-to-unsprung mass was $\rho = 10$, and the unsprung natural frequency was $\omega_{us} = 10 \times 2\pi$ rad/s. The PI weights r_1 and r_2 serve as "tuning" knobs controlling the optimal trade-off between acceleration, suspension stroke (design constraint), and tire deflection (handling constraint).

From Figure 31.9b it follows that reducing r_1 while keeping r_2 constant results in increased tire deflection and decreased sprung mass acceleration. On the other hand, Figure 31.9a shows that reduction of r_2 while keeping r_1 constant leads to increased suspension deflection and decreased accelerations. In the extreme case of $r_2 = 0$, the problem becomes a degenerate one where extremely large suspension deflections are possible since the suspension force is not a function of suspension deflection and thus cannot give static support for the vehicle weight. Indeed, Figure 31.9 shows that there is no incentive to increase r_1/r_2 above ~1000, since then the acceleration and tire deflection remain practically constant, while the suspension deflection increases significantly. Similar problems do not occur if only $r_1 = 0$, and this case can be used for preliminary suspension design. As a rule of thumb, $r_1/r_2 \sim 10$ leads to reasonable results. In addition, the area of relatively large normalized accelerations (above say 50) in Figure 31.9a and b corresponds to very large disturbance levels typical for off-road operations, whereas the area of small normalized acclerations is characteristic of very smooth roads and/or low vehicle speeds.

When these considerations are taken into account, it becomes evident that most optimal designs of practical interest for automotive applications will fall within the shaded area in Figure 31.9a and b, with normalized accelerations between 10 and 50 $s^{-3/2}$. This is consistent with corresponding passive results discussed in Section 31.5.2.2 later. Also, based on most vehicle/road data for on-road operations, it is to be expected that the tire deflection constraint will trigger first, that is, become active before the rattlespace does. This means that in most cases, the graph from Figure 31.9b should be checked first, and when the tire handling constraint there is satisfied, there is a good chance that the rattlespace constraint will be satisfied too. Finally, note that no control design, linear or nonlinear, can fall below the $r_1 = 10^{-2} \sim 0$ line in Figure 31.9a and $r_2 = 10^{-3} \sim 0$ line in Figure 31.9b. Typical usage of the global performance plots is best shown through an illustrative example.

Similar solutions can also be obtained using linear matrix inequality (LMI)-based approaches [52–54], although one should be aware of occasional incorrect claims that this type of methods outperform the earlier carpet plots in their native H_2 metric. A desirable feature of the LMI-type controllers is that they can incorporate the (slowly) time-varying parameters, such as changes in vehicle sprung mass. Related methods here include the linear parameter-varying (LPV) techniques [53,55]

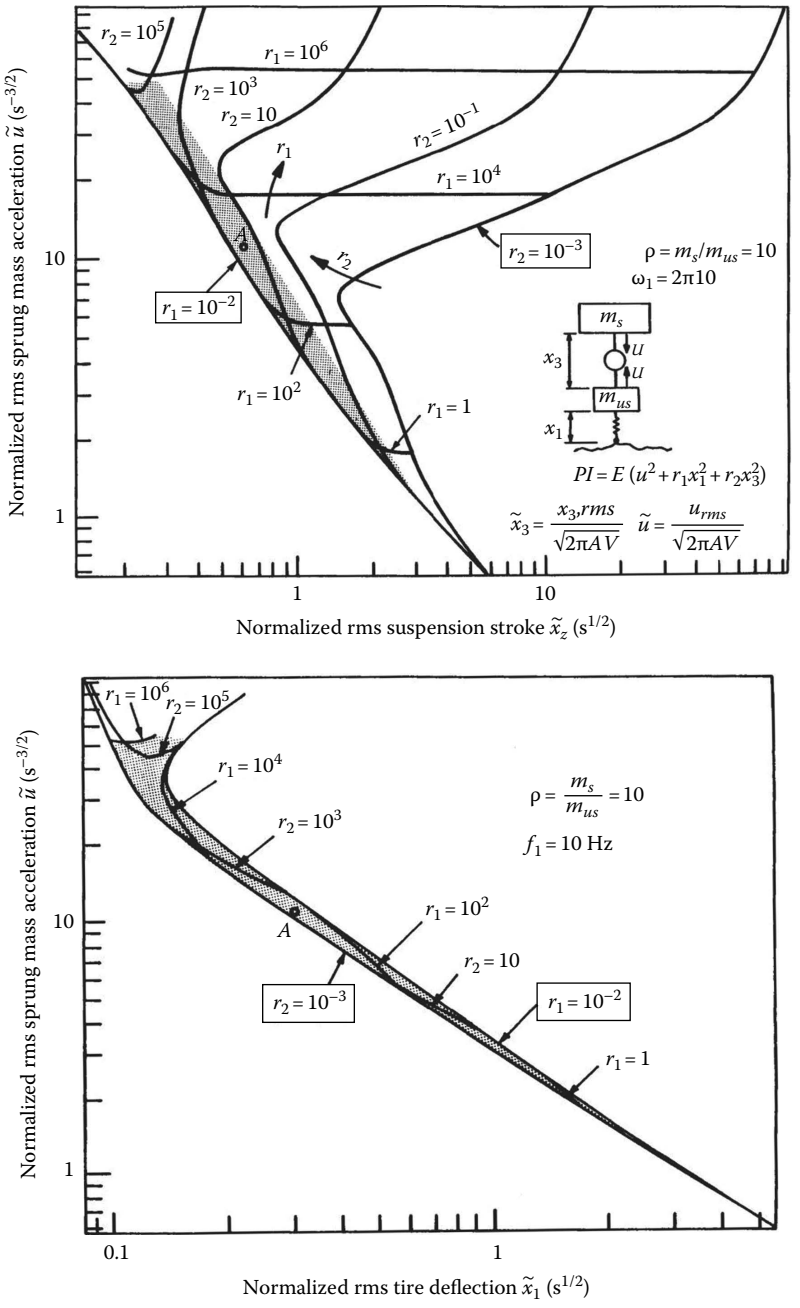


FIGURE 31.9 (a) Optimal normalized rms acceleration versus normalized rms suspension stroke for 2-DOF model. (b) Optimal normalized rms acceleration versus normalized rms tire deflection for 2-DOF model.

that explicitly include possible slowly varying system parameters, such as road roughness and vehicle speed variations with results similar to [56]. In general, there are many publications that essentially repeat original results obtained in the 1980s and early 1990s [24] but with different techniques and approaches, which could be attributed to the "waterbed" effect where any improvement in one area (e.g., ride) is reflected through corresponding loss of performance in other, complementary areas (e.g., handling and/or design/rattlespace requirements).

31.5.2.1 Illustrative Example

Assume that for a particular vehicle $\rho = 10$ and $\omega_{us} = 2\pi \times 10$ rad/s. The road is described by a road roughness coefficient $A = 1.6 \times 10^{-5}$ ft (4.9×10^{-6} m) that corresponds to a medium-quality road. The vehicle is traversing this road at a speed of $V = 80$ ft/s (88.5 km/h). The tire deflection from equilibrium should be below 1 inch (2.54 cm) 99.7% of the time so that x_{1rms}<1/3 in. = 0.847 cm and $\tilde{x}_1 = x_{1,rms}/\sqrt{2\pi AV} < 0.3\ s^{1/2}$. From Figure 31.9b the minimum possible normalized rms acceleration is $\tilde{u} = 10\ s^{-3/2} = 10\ s^{-3/2}$. Accepting $\tilde{u} = 10.9\ s^{-3/2}$ results in 3% g rms acceleration and tuning parameters $r_1 = 1100$ and $r_2 = 100$ shown as design A. Using these values for r_1 and r_2 in Figure 31.9a reveals a normalized rms secondary suspension deflection of $x_3 = 0.605\ s^{1/2}$ that insures that the suspension deflection will remain within ± 2 in. (5.08 cm) of the static value 99.7% of the time. With $r_1 = 1000$ and $r_2 = 100$ the optimal control gains are $k_1 = -6.084$, $k_2 = 0.548$, $k_3 = -10.0$, and $k_4 = -4.438$, and the closed-loop eigenvalues are $e_{1,2} = -2.20 \pm j2.26$ and $e_{3,4} = -2.75 \pm j62.9$. The first set of oscillatory eigenvalues corresponds to the vehicle heave mode with a natural frequency of 0.5 Hz and damping ratio of ≈0.7. The second set corresponds to the wheel-hop mode with a resonant frequency of 10 Hz and relatively small damping ratio of 4.4%. Whether or not this small amount of damping is of consequence will depend on the particular hardware implementation, which may include a combination of active and passive means to improve closed-loop robustness. Also, in practice, an excessively soft suspension setting may require load leveling to keep the suspension static or steady-state deflection in the neighborhood of zero or nominal value.

31.5.2.2 Comparison with Conventional Passive Suspension

At this stage, it is instructive to compare the earlier active suspension results with corresponding passive case. For this purpose consider Figure 31.10 from Hrovat [24], which displays normalized rms acceleration versus tire deflection for active and passive suspensions. For simplicity, the optimal active suspension performance is shown via the limiting curves $r_1 \sim 0$ and $r_2 \sim 0$ only, which are representative of the main trends. Here, the design points with lower accelerations and higher tire deflections are characterized by softer, lower-frequency body modes and less damped wheel-hop modes.

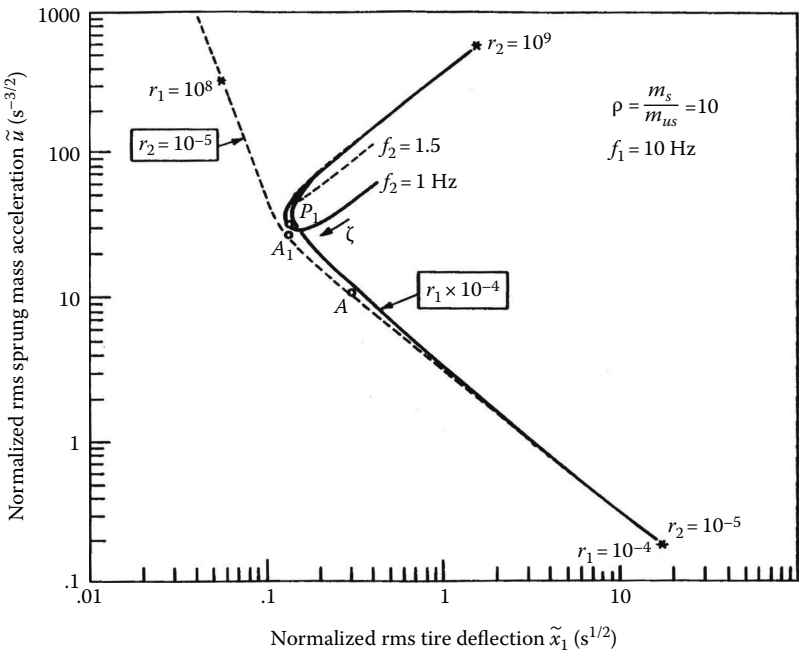


FIGURE 31.10 Comparison between conventional passive suspension and optimal active suspension.

The passive performance trade-offs correspond to vehicles characterized by a structure shown in Figure 31.8c, with heave mode resonant frequencies f_2 between 1 and 1.5 Hz, and damping ratios, ζ, between 0.02 and 1. A typical passive suspension setting is shown as point P_1, where $f_2 = 1$ Hz and $\zeta = 0.3$. As can be seen from Figure 31.10, this amount of damping is near optimal in terms of ride and handling compromise for the passive case. These values are within the range typically found in most of today's vehicles, which have evolved through many iterations primarily based on intuition and decades of experience.

The performance of an optimal active suspension with the *same* amount of wheel hop is shown in Figure 31.10 as point A_1. Comparing P_1 and A_1, it can be seen that in this case the active suspension results in only 11% lower rms acceleration levels. However, it can also be seen that for some other road/speed operating conditions (to the right of A_1), there is a potential for substantial ride improvement with active suspensions compared with their passive counterparts. Indeed, as shown in Figure 31.10, for passive suspensions almost any deviation in operating conditions from P_1 will result in degradation of performance. On the other hand, for active suspensions, either handling or ride can be improved by choosing the tuning parameters r_1 and r_2 such that the resulting operation settles either to the left or to the right of A_1, respectively. For example, by relaxing the wheel-hop requirements, the acceleration for design case A treated in the earlier illustrative example is 67% smaller than the one at P_1—see Figure 31.10.

Different attributes of the performance at three operating points of interest (A, A_1, and P_1) can also be seen from associated Bode plots shown in Figures 31.11a through c. From Figure 31.11c, it can be seen that the main difference between the active suspension A_1 and corresponding passive counterpart P_1 occurs at the sprung mass mode around 1 Hz, where the active suspension brings much more damping through the aforementioned "skyhook" effect. The active suspension A results in further significant reduction of the acceleration levels (except at the wheel-hop invariant point of 10 Hz—see 31.5.2.5 in the following text), but at the same time it deteriorates handling and rattlespace requirements due to a large, poorly damped resonant peak around the wheel-hop frequency around 10 Hz. It can also be shown [24] that, unlike their passive counterparts, the LQ-optimal suspensions can result in a nonzero DC gain between the rattlespace and road velocity input. This may not be too detrimental in practice, since it is possible to use high-pass filters to eliminate the very low-frequency components of the ground input signals resulting from hills and extended road grades.

Based on the earlier results and observations, it can be concluded that the full advantage of active suspensions stems from possible *adaptive* tuning (or gain scheduling) of controller parameters, depending on the driving condition [24,56]. For example, if the steering wheel position or lateral acceleration sensors indicate operation on a straight section of a road, where handling is less critical, then it may be possible to relax the wheel-hop constraint. The exact amount of wheel hop that could be tolerated under different driving conditions should be predetermined through appropriate vehicle tests.

31.5.2.3 Comparison with 1 DOF

Let us now place the earlier 2-DOF, quarter-car results in proper perspective with respect to their 1-DOF counterparts. For this purpose we will first introduce a new, modified 1-DOF model, which will facilitate comparisons with 2-DOF systems. The new 1-DOF model, shown in Figure 31.8b, corresponds to a special, limiting 2-DOF case, where the unsprung mass is reduced to zero (note that, as shown in [57], the optimal performance improves as the unsprung mass decreases (see 31.5.2.4 in the following text)). The optimal suspension control problem for this special case has been analytically solved with the help of appropriately defined states [57]. The resulting optimal trade-off curves are shown in Figure 31.12, for the limiting cases when $r_1 = 10^{-4}$ and $r_2 = 10^{-5}$ in the PI of Equation 31.9.

Note from Figure 31.12a that for normalized rms suspension strokes, $\tilde{x}_3$, above roughly 0.3, this new 1-DOF performance is similar to the traditional one. However, for $\tilde{x}_3$ below 0.3 the new 1-DOF

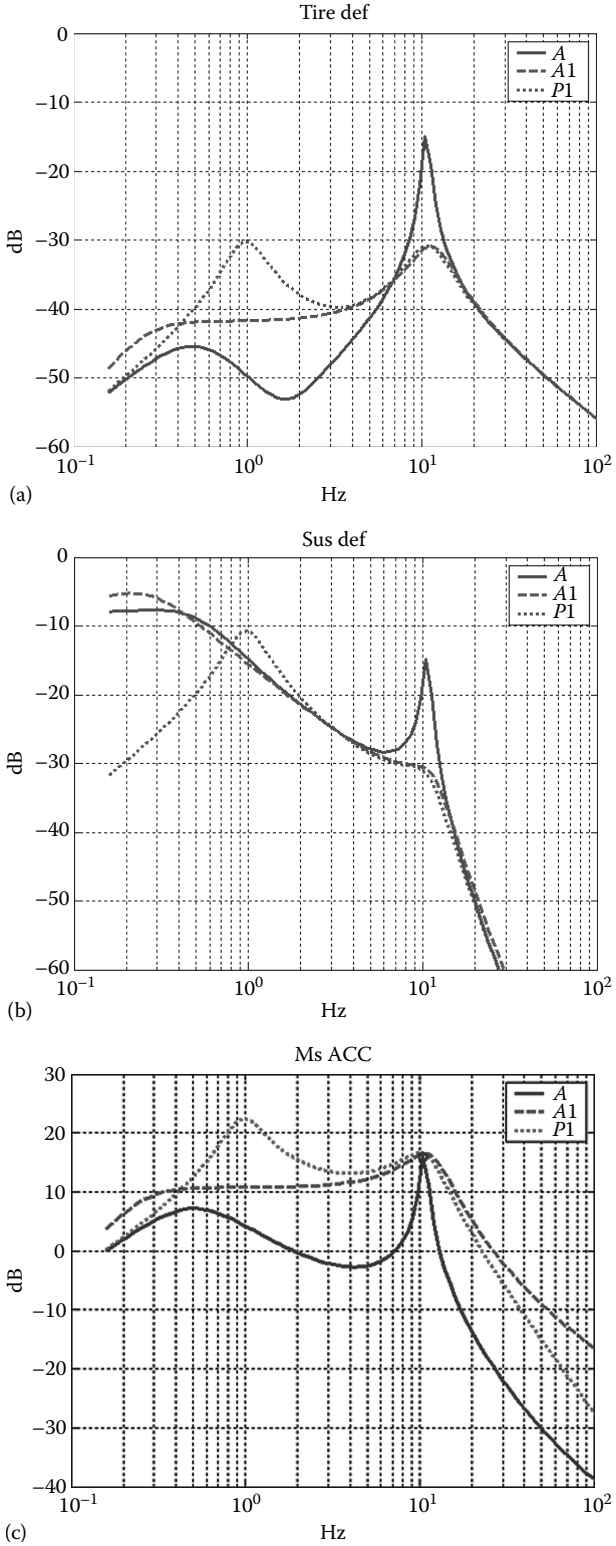


FIGURE 31.11 (a) Bode plots of tire deflection, (b) rattlespace, (c) sprung mass acceleration versus ground velocity for associated operating points A, A_1, and P of Figure 31.10.

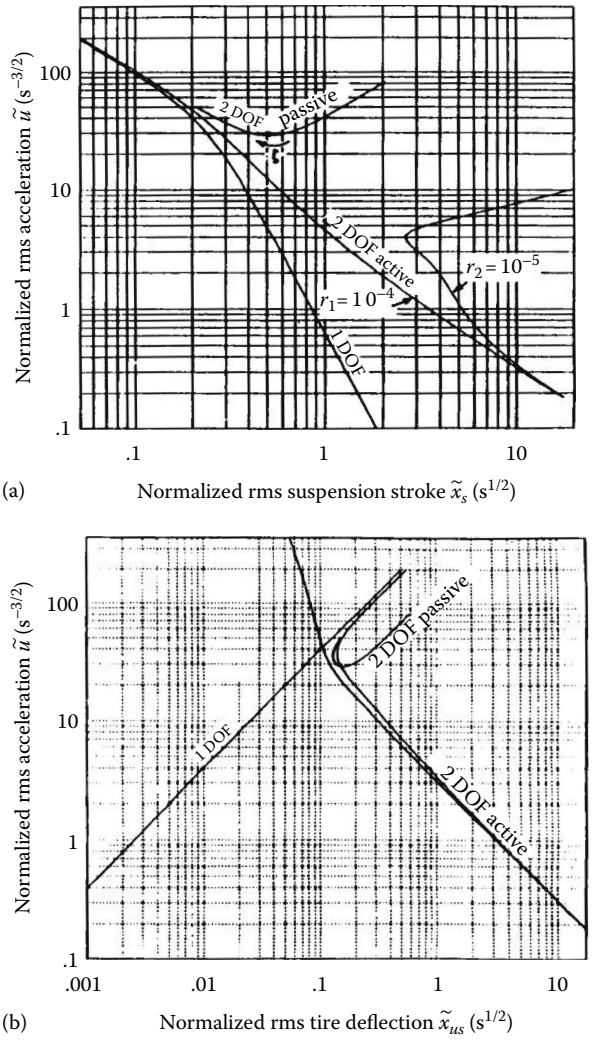


FIGURE 31.12 Comparison, in particular, (a) suspension stroke and (b) tire deflection, between conventional passive suspension, optimal 2-DOF suspension, and optimal limiting case where *mus* = 0.

curve is deflected toward smaller acceleration values. This is due to the gradual stiffening of the (secondary) suspension, so that more and more of the overall deflection is now being taken by the (primary) tire spring. Of course, the overall deflection is still larger than in the traditional optimal 1-DOF case.

Also displayed in Figure 31.12 are the corresponding limiting performance curves for typical 2-DOF passive and optimal active suspensions. It can be seen that while the optimal active suspensions for 2-DOF systems can still significantly outperform passive ones, they fall short of the optimal 1-DOF performance. For example, at a normalized rms suspension stroke of 0.6 $s^{1/2}$, the 1-DOF case results in a normalized rms acceleration of 3 $s^{-3/2}$, whereas the 2-DOF results in 10.9 $s^{-3/2}$, a 263% increase. The main reason for this deterioration stems from the conflicting requirements imposed on the active actuator of Figure 31.8a: it should simultaneously provide small sprung mass acceleration (for comfortable ride) and a considerable amount of unsprung mass damping needed to reduce the wheel hop (for good handling). It is easier to satisfy these conflicting requirements for smaller unsprung mass, which gives an additional incentive for reducing the unsprung weight through the use of, for example, aluminum wheels and lightweight, composite materials. The best

performance within the present single-actuator 2-DOF structure of Figure 31.8a is then obtained when $m_{us} = 0$, as shown in Figures 31.12a and b.

31.5.2.4 Influence of Unsprung Mass

The preceding discussions already indicated that in the context of the 2-DOF quarter-car problem, it is advantageous to reduce the unsprung mass or—more specifically—to reduce the ratio of the unsprung-to-sprung masses. A more detailed study to that effect has been documented in [57]. One of the main conclusions of this study is that for the case when the total vehicle mass is kept constant, every 10% reduction in unsprung mass contributes to 6% reduction in rms sprung mass acceleration, for the same level of wheel hop. For active suspension vehicles, this gives an added incentive for reduction of unsprung weight through the usage of lightweight materials and appropriate design changes.

31.5.2.5 Additional Characteristics: Invariant Points

Using the earlier 2-DOF case, it was possible to point out the limitations imposed by introduction of the additional DOF and associated handling (i.e., tire deflection) constraint. Additional insight can be derived from the fact that for certain transmission paths, there are frequencies where the associated transfer functions are constant or invariant. One such invariant point exists at the wheel-hop resonance of

$$\omega_1 = \sqrt{\frac{k_{us}}{m_{us}}} \tag{31.11}$$

in the case of a transfer function between sprung mass acceleration and road input velocity [7]. This can be seen from Figure 31.11c, which illustrates that in the immediate vicinity of the wheel-hop resonant frequency, all acceleration transfer function levels are about the same, independent of the suspension type (active or passive) and parameters. One additional invariant point [58] exists for the transfer function between rattlespace and road input velocity at the intermediate frequency

$$\omega_2 = \sqrt{\frac{k_{us}}{(m_s + m_{us})}} \tag{31.12}$$

which corresponds to the case when the two masses, m_s and m_{us}, are riding locked together on the tire with stiffness k_{us}. This invariant point can be seen on Bode plots in Figure 31.11b. It should be pointed out that introduction of tire damping eliminates the earlier invariants [24], although, since in practice the amount of tire damping is typically small, the knowledge about the earlier invariant points is useful and insightful information.

31.5.2.6 Dynamic Vibration Absorbers

The previous two sections show some fundamental limitations of active suspensions when applied to a standard 2-DOF, quarter-car structure. Chiefly, one of the main constraints of such a structure is that one and the same suspension force must address two conflicting requirements: on one hand, it must be small enough to provide for a comfortable ride, while at the same time it should also be large enough to provide adequate wheel-hop damping for improved vehicle handling. This limits maximum achievable performance for the given structure, so that any further improvements are only possible through *fundamental structural* or design changes. One possible solution here can be found in the form of a dynamic vibration absorber or tuned mass damper [24,59], attached to the unsprung mass as shown in Figure 31.13.

Dynamic absorbers (DAs) are often used to contain pronounced and lightly damped oscillations. Typical applications of DA range from electric hair clippers [60] to tuned mass dampers

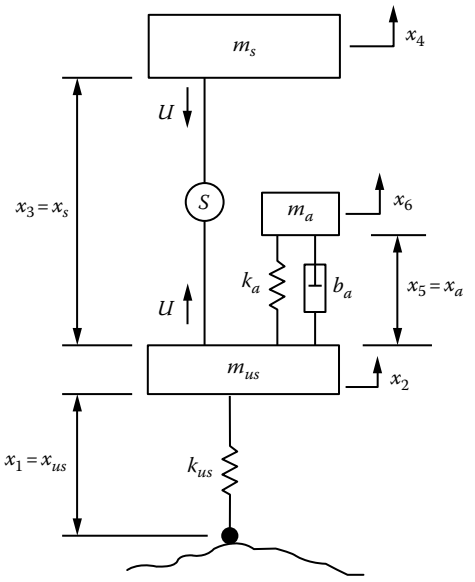


FIGURE 31.13 Two-DOF vehicle model with DA.

for structural control of tall buildings [61]. In automotive applications, DAs have often been used, sometimes as afterthoughts, to reduce unforeseen and undesirable vibrations. However, apparently there has been only one production application of DA with the aim of improving ride comfort. This is the case of the well-known French subcompact Citroen 2CV, first introduced in 1949. The 2CV was legendary for having superior ride for a car of its size.

A global study of the potential benefits of DA in the context of active suspensions was reported in [59]. Both passive and active DA concepts have been considered in this study. The resulting plots of limiting performance trade-offs with an added passive DA are shown in Figure 31.14, where the traces marked "2 DOF + DA" correspond to optimal 2-DOF controller gains with an appended DA with the mass equal to 10% of the unsprung mass, natural frequency $\omega_a = \omega_{us} = 2\pi \times 10$ r/s, and a damping ratio $\zeta = 0.2$. Comparing the 2 DOF + DA case with the optimal 2-DOF and 1-DOF trade-offs in Figure 31.14a, it can be seen that the DA can contribute to substantial improvements in performance, especially for $\tilde{x}_s > 0.2$ s$^{1/2}$.

As an example, consider again operations at point *A* from Figures 31.9 and 31.10, which had an unsprung mass damping ratio of only 4.4%. As indicated earlier in 31.5.2.1, this small amount of damping can be inadequate, and some way of increasing it would be highly desirable. Here we propose to use an appropriately tuned DA. The addition of such DA results in a performance at A′ in Figure 31.14, where acceleration, rattlespace, and tire deflection have all been reduced, while at the same time the wheel-hop damping ratio increased to 16%. Further substantial reduction in acceleration results for point *A″*, which has the same suspension stroke requirements and lower tire deflection than *A*.

This clearly illustrates that the addition of DA has significant potential for improved ride and handling performance of active suspensions. Their main drawback is the added weight and packaging requirements, which represent a real design and development challenge that must be resolved prior to their possible widespread usage. One step in this direction might be the recently advertised Bose active suspension [62]. It seems to incorporate the DA (or so-called damper mass [62]) integrated in the wheel assembly in an apparently efficient way [63]. Additional vehicle objective and subjective tests and evaluations are needed to fully ascertain the effectiveness of this particular suspension. Finally, further performance benefits are possible with the help of various forms of active DAs and unsprung mass actuators as investigated in [59].

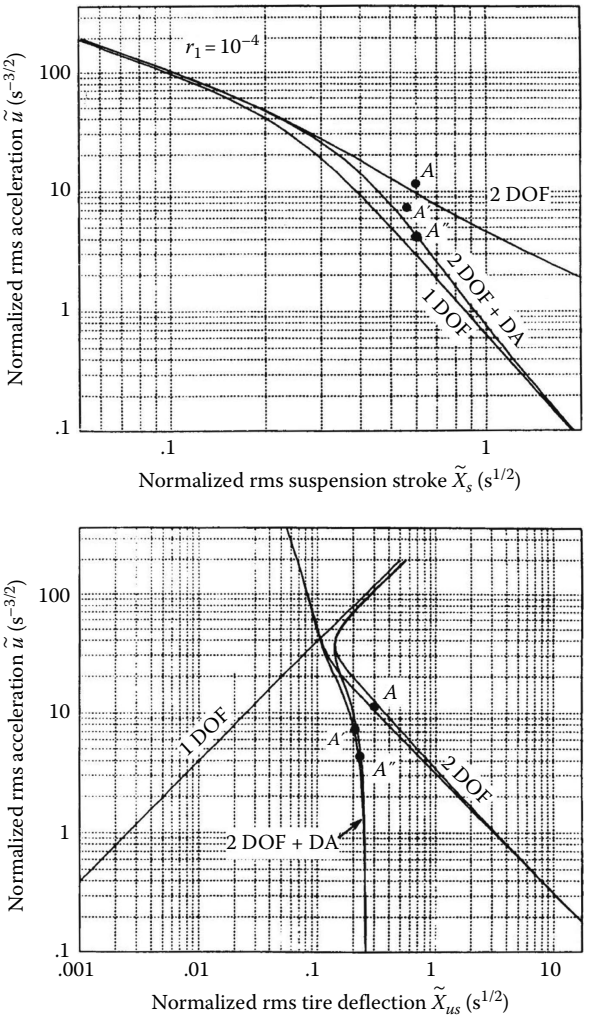


FIGURE 31.14 Comparison between optimal 2-DOF suspension ($\rho = 10$, $\omega_{us} = 2\pi \cdot 10$ s^{-1}), a suboptimal 2 DOF + DA suspension ($\rho_a = \rho$, $\omega_a = \omega_{us}$, $\varsigma_a = 0.2$), and optimal limiting case when $m_{us} = 0$, $r_1 = 10^{-4}$, and $r_2 = 10^{-5}$.

31.5.2.7 Optimal Quarter-Car Structure

In the preceding section, we have seen that changing the quarter-car structure via (tuned) DA can substantially improve the performance of active suspension vehicles. At this point, it is appropriate to ask what is the best possible quarter-car structure that would result in an absolute minimum for PI_3 of Equation 31.9. For this purpose, we introduce the suspension configuration from Hrovat [59], shown in Figure 31.15a. Unlike the conventional 2-DOF structure from Figure 31.8a, the proposed configuration is not constrained by the requirement of equal sprung and unsprung forces, U_s and U_{us}. This implies that such a two-control configuration is very flexible and powerful—sort of a "dream team" of quarter-car suspension structures. For example, the unsprung control force U_{us} can be used to alter both the effective unsprung mass and tire stiffness and damping without negatively affecting the sprung mass acceleration that is controlled via the sprung mass force U_s. While this is conceptually simple and elegant, practical implementation of this intriguing "dream team" concept is by no means trivial. For start, it is practically not possible to find an inertial ground from a moving vehicle. (Similar effect can be realized via reaction jets, but this, as of

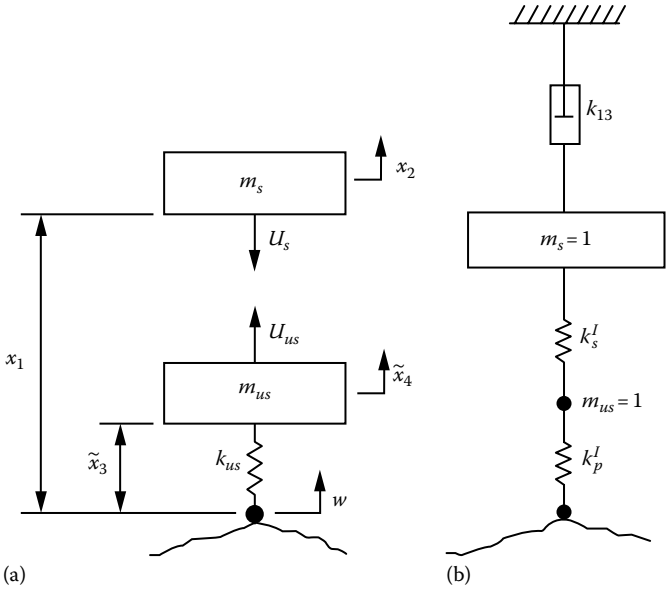


FIGURE 31.15 (a) Two-DOF vehicle model with active suspension capable of producing independent forces U_s and U_{us}. (b) Singular optimal incremental suspensions structure.

now, is not practical.) An approximation can be obtained via active dynamic vibration absorbers discussed in the previous section. Another practical consideration relates to the required unsprung force production bandwidth, which can be quite large and challenging.

As shown by Hrovat [59], the earlier inquiry into the best possible 2-DOF quarter-car performance leads to a singular optimal control problem characterized by a lack of penalty on the unsprung actuator force. This is known as a "partially cheap optimal control problem." In this context the cheap, unsprung control primarily serves as a structure optimizer: it can effectively reduce or eliminate the unsprung inertia and add or substitute additional inertia, compliance, and damping as needed to minimize PI_3.

The solution to the earlier optimization problem was obtained in [59] by transforming the aforementioned cheap control problem into an equivalent singular perturbation setting, as discussed in [24] and references therein. The resulting optimal structure is shown in Figure 31.15b. With the hindsight from Hrovat [59], this simple structure can be explained as follows: The sprung mass control U_s is used to contain the slow body-related mode in accordance with an appropriately modified, "standard" 1-DOF optimal control law with the damping ratio of 0.7. The cheap and fast control U_{us} accomplishes two tasks. In effect, it completely eliminates the unsprung mass and adjusts the primary and secondary incremental stiffnesses so that the unsprung motion can instantaneously adapt to the best possible configuration for a given set of performance weights, r_1 and r_2.

The presence of two independent control forces, U_s and U_{us}, causes the new 1-DOF optimal structure of Figure 31.15b to substantially differ from the previously discussed 1-DOF configuration of Figure 31.8b, since the latter cannot alter the primary suspension stiffness. As such the new 1-DOF optimal structure serves as a benchmark of the best possible PI_3 performance that can be obtained with 2-DOF, quarter-car models. For an illustrative example introduced in [59], this limiting case resulted in normalized rms sprung mass acceleration of only 1.17 $s^{-3/2}$, with normalized rms tire and suspension deflections of 0.23 and 0.6 $s^{1/2}$, respectively (to fully appreciate this performance, consider Figure 31.14). This example illustrates that further substantial ride and handling improvements are possible with high-performance, active, and SA unsprung mass actuators. While in some cases this superior performance may not be practical due to packaging, weight, and cost constraints, it serves as an absolute reference against which any practical implementation can be compared.

31.5.2.8 Additional Considerations

The 2-DOF models used thus far have been developed under the assumption of zero tire or wheel-hop mode damping. This is a standard assumption, supported by the fact that the unsprung damping is usually very small. However, it is of interest to mention that the results of a global numerical study [57] demonstrated that even a tire damping ratio of only 2% can significantly improve the shape of the limiting optimal curves, which become much closer to the optimal 1-DOF curves in Figure 31.14. It was also shown there that even in this case, the DA still offers tangible benefits, which are, of course, somewhat less dramatic than before. Since in practice the tire damping can vary from a few percent at low speeds to near zero at high speeds, we should have the earlier results in mind when evaluating different passive and active suspension concepts.

It should be noted that, due to a natural separation between the slow body mode and the fast wheel-hop mode, it is possible to use *singular perturbation* theory to partition the problem into an equivalent slow 1-DOF system and a fast wheel-hop-associated subsystem [64,65]. This also creates a link with and further justification for 1-DOF studies.

In typical practical implementations, there will already be some sort of passive suspension in place to minimize active energy requirements and secure "limp-home" capability in case of active actuator failure. The active suspension then acts in tandem with its passive counterpart and as such may also enhance the overall system robustness since the added structural damping results in better damped zeros for the associated force-tracking task (see Section 22.3.3). Reference [66] considers simultaneous optimization of both active force and passive suspension parameters. A PI similar to Equation 31.9 was used with an additional term proportional to the square of active actuator force as a rough approximation of the corresponding energy requirements (e.g., in the case of electrohydraulic actuators without regeneration capabilities). The corresponding simulation results show potential for additional benefits especially in terms of energy consumption as reflected in this expended PI metric.

Finally, one could parallel the 1-DOF case and introduce *jerk* as an additional ride term in PI_3, Equation 31.9. For an overview of related applications for 2-DOF, quarter-car systems, consult [24] and references therein. In the case of 1-DOF systems, the jerk-optimal LQR solution lead to in practice very attractive structure composed of a skyhook damper and spring and *an integrator* operating on rattlespace error (deviation from equilibrium) in the manner of a fast load leveler. A similar integrator could be a priori imposed to the corresponding 2-DOF case leading to an LQ-optimal solution. As in the more conventional 2-DOF case, the jerk-augmented performance can be further improved via tuned dynamic vibration absorbers and other structural means—see [24] and references therein.

In summary, the optimal active suspensions for 2-DOF quarter-car models without tire damping result in significant loss of performance with respect to their 1-DOF counterparts. Nevertheless, such an optimal 2-DOF system can still substantially outperform the conventional passive suspensions due to the *inherent adaptive* capabilities of active suspensions, which can adapt to different road roughness levels and vehicle operating conditions. Further improvements are possible through a reduction of the unsprung mass and through structural changes aimed at independently controlling the wheel-hop mode in a manner that would not negatively impact the quality of the isolation of the sprung mass vibration. Ultimate quarter-car suspension would include an active unsprung mass actuator that would result in the most favorable structure and superior performance.

31.6 OPTIMAL SUSPENSIONS FOR HALF-CAR MODELS

There are a number of publications (e.g., [67,68]) investigating optimal and suboptimal ride characteristics of half-car, 2D vehicle models, which include both sprung mass heave and pitch modes as shown in Figure 31.16. Most of the past work was based on numerical analysis, which was used to evaluate optimal active suspension performance at a (usually) single operational point.

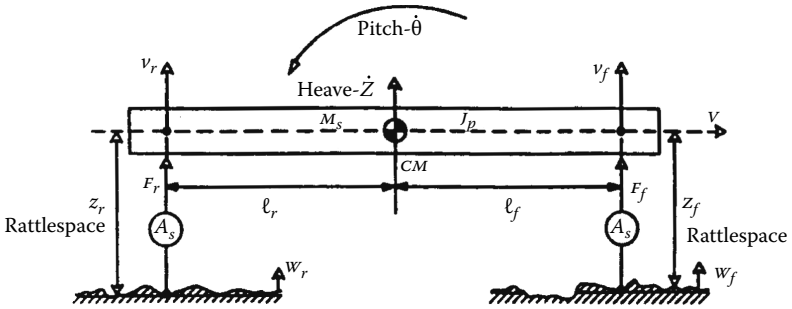


FIGURE 31.16 Half-car 2-DOF model.

This was extended in Krtolica and Hrovat [69], which developed a global analytical solution for the optimal suspension control problem minimizing

$$PI_4 = E\left(r_1\ddot{z}^2 + r_2\ddot{\theta}^2 + r_3 z_f{}^2 + r_4 z_r{}^2\right) \tag{31.13}$$

subject to 2-DOF, fourth-order system state equations describing vehicle heave and pitch dynamics as per Figure 31.16. In Equation 31.13 the first two terms penalize excessive heave and pitch accelerations, $\ddot{z}$ and $\ddot{\theta}$, respectively, while the last two terms penalize excessive front and rear suspension deflections, z_f and z_r.

As a by-product of the analytical solution, it was possible to establish a number of properties characterizing the resulting LQ-optimal suspension. For example, as in the simple 1-DOF quarter-car case discussed in Section 31.5.1, the optimal half-car suspension is also characterized by a favorable damping ratio of 0.7. The preceding reference by Krtolica and Hrovat [69] also contains necessary and sufficient conditions for decoupling of the original 2-DOF, half-car problem into two 1-DOF, quarter-car subproblems. These conditions are

$$M_s \cdot l_f \cdot l_r = J_p \tag{31.14}$$

$$r_1 \cdot l_f \cdot l_r = r_2 \tag{31.15}$$

where

M_s and J_p are vehicle mass and the pitch moment of inertia about the CM
l_f and l_r are the front and rear distances from the CM (cf. Figure 31.16)

The first condition, Equation 31.11, depends on vehicle physical parameters and is approximately (within 20%) satisfied by most vehicles. The second condition, Equation 31.12, can be satisfied by appropriately choosing the weighting parameters r_1 and r_2, which often results in a reasonable compromise between the pitch and heave aspects of a ride.

The same conditions also apply for the 4-DOF, half-car model that includes unsprung masses as per Figure 31.17a. The result is two decoupled 2-DOF, quarter-car models shown in Figure 31.17b. Note that the 2D, half-car model introduces one qualitatively new aspect—consideration of the front wheels as a convenient mechanism to preview the road ahead of the rear wheels. This anticipatory information can further improve the ride/handling trade-offs, above and beyond the optimal quarter-car trade-offs of Figures 31.9a and b.

Global performance potentials with preview applied to simple 1-DOF quarter-car models were investigated in Bender [4] and Tomizuka [70]. Hrovat [71] introduces an equivalent delayed-system formulation to develop similar results for 2-DOF, quarter-car, discrete-time models applicable to the configuration of Figure 31.17b. The related global performance plots are reproduced in Figure 31.18, where the solid lines represent the limiting curves without the preview. The optimal performance curves with preview times t_r up to 1 s are shown as dashed lines.

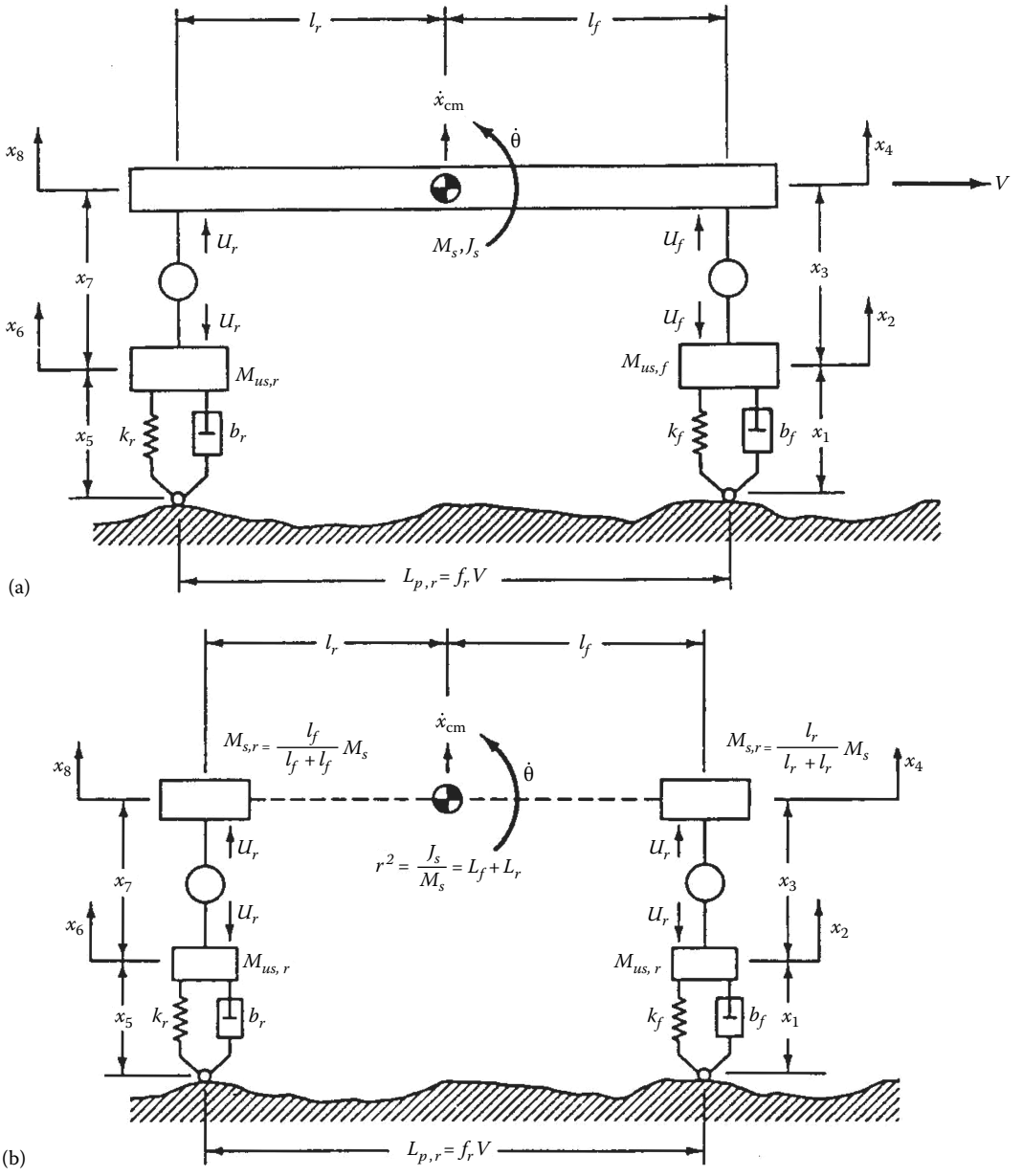


FIGURE 31.17 (a) Half-car, 4-DOF vehicle model. (b) Corresponding decoupled model consisting of two quarter-car 2-DOF subsystems.

From Figure 31.18a it can be seen that for normalized suspension deflections below about 0.6 $s^{1/2}$, preview information of up to 1 s can reduce the accelerations by approximately 50%–70%. Similarly, Figure 31.18b shows that even a relatively small amount of preview of only 50 ms can significantly reduce the tire deflection for normalized accelerations above circa 15 $s^{-3/2}$. For example, at normalized accelerations of 20 $s^{-3/2}$, the tire deflection can be reduced by about 30% with 50 ms of preview. Since in practice the most useful previews will probably be limited to below about 0.3 s, it can be concluded that preview information will probably be most useful for normalized accelerations above approximately 10 $s^{-3/2}$.

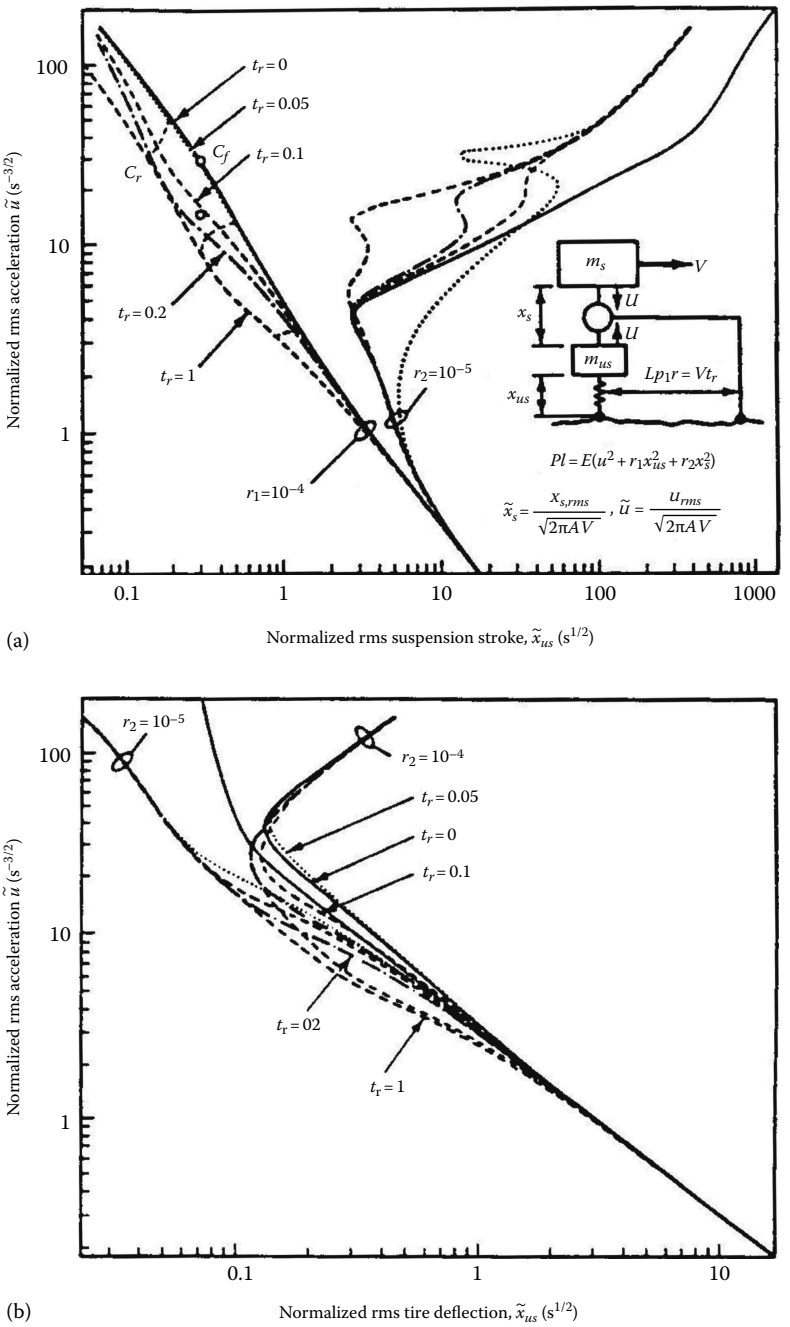


FIGURE 31.18 Optimal suspension performance for 2-DOF model with different preview times t_r, and ρ = 10, $\omega_{us} = 2\pi \cdot 10$ s^{-1}.

Moreover, one can now use the earlier plots to adapt different control gains for the front and rear suspensions, while satisfying tire and suspension deflection constraints. This is illustrated in Figure 31.18a with the help of design points C_f and C_r, for the front and rear control settings. Here both suspension units are subjected to the same stroke constraint, while the rear unit can benefit from 0.11 s of preview, which corresponds to a vehicle with a wheelbase of 2.75 m traversing a fair-to-medium road at a speed of 88.5 kph (further details about this example can be found in [71]). The result is a

considerable improvement in both heave and pitch aspects of the ride. Obviously, if a suitable sensor were available, then the performance could be further improved with the road preview ahead of the front wheels. Corresponding additional benefits could be again estimated with the help of Figure 31.18.

As in the quarter-car case, it is now possible to consider jerk-optimal suspensions for half-car models [72] and, also, the usage of singular perturbation theory to exploit the two-time scale separation between body and wheel-hop modes [64,73]. In addition, with the half-car structure it is possible to exploit the spatial separation between the front and rear of the vehicle. This was first done in [64], where the authors essentially neglected the cross-coupling between the front and rear PI terms corresponding to the fast, wheel-hop dynamics. The resulting hierarchical control structure consisted of a slow mode similar to the one from Figure 31.16 (with a set of tire springs as per Figure 31.8b) and a decoupled set of fast control modes used to contain the front and rear wheel hops. However, as shown by [65], for the particular data used in [64], the decoupling conditions (31.11) and (31.12) are almost satisfied anyway, so that even a simple set of two fully decoupled front and rear quarter-car models yields a performance close to optimal. It is only when strongly coupled data are used that one can start to appreciate the proposed spatial and temporal separation.

A comprehensive analysis of the 2D preview case in the continuous-time domain was performed by Hac [74]; the corresponding analytical solutions were similar to their limiting discrete-time counterparts from [71]. The performance potentials of front wheel preview have been further investigated in [75] and those of the slow-active suspensions in [76,77], where it was concluded that "0.033 s of preview is of little use, although 0.1 s is very useful." An attempt to apply the preview to a nonlinear 2D truck model has been discussed in [78], along with the corresponding power requirements. Finally, an analytical application of decentralized control concepts to active vehicle suspensions with preview is given in [79].

31.7 OPTIMAL SUSPENSIONS FOR FULL-CAR MODELS

In this section we will briefly survey some past and recent full-car, 3D results and place them in a proper perspective with respect to previously discussed half-car and quarter-car counterparts. Some of the earlier investigations of the optimal 3D suspensions can be found in [80–82], where numerical solutions of the corresponding LQ problem were obtained for a few specific operational road/speed conditions. On the other hand, complete analytical solution of the 2D case [69] facilitated similar analytical extensions [83] toward the full 3D structure shown in Figure 31.19. The corresponding

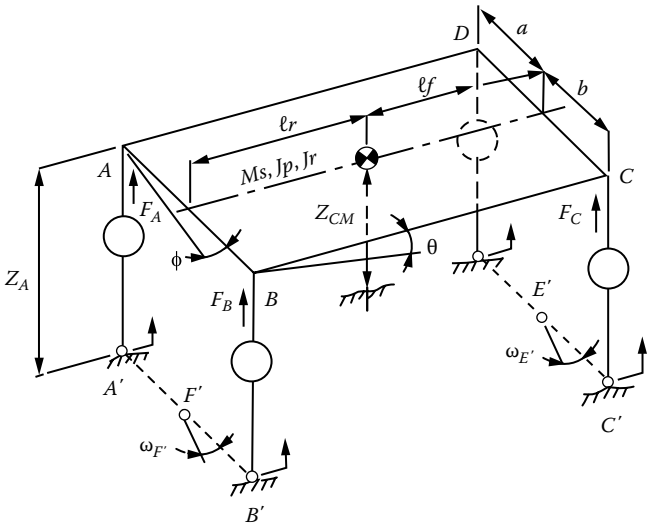


FIGURE 31.19 A 3D-vehicle model schematic.

optimization problem was formulated by mimicking the basic quarter-car and half-car performance requirements, Equations 31.8 and 31.9, so that the full-car PI now becomes

$$PI_5 = E\left(q_A z_A{}^2 + q_B z_B{}^2 + q_C z_C{}^2 + q_D z_D{}^2 + r_1 \ddot{z}_{cm}{}^2 + r_2 \ddot{\Theta}^2 + r_3 \ddot{\Phi}^2\right) \tag{31.16}$$

where the first four terms penalize excessive steady-state suspension deflection at the four corners (cf. Figure 31.19), while the last three serve as a measure of ride discomfort with respect to vehicle heave, pitch, and roll acceleration, respectively.

The analytical solution of the earlier seventh-order system was found with the help of a two-layered partitioning [83]. In the first step, the original seventh-order system was partitioned into a sixth-order, tricycle structure shown in Figure 31.20 and an additional "odd" state. The second step constituted further partitioning of the sixth-order Riccati equation, corresponding to the earlier tricycle model, into two third-order subsystems, which could then be solved analytically. The analytical solution of the sixth-order system was next merged with the additional seventh state shown in Figure 31.21. This additional state reflects the fact that the road elevations are not necessarily coplanar at the four tire–road interfaces.

The solution to the seventh-order LQ problem was completed by transforming the earlier seven states into a more practical set of states: three sprung mass velocities for heave, pitch, and

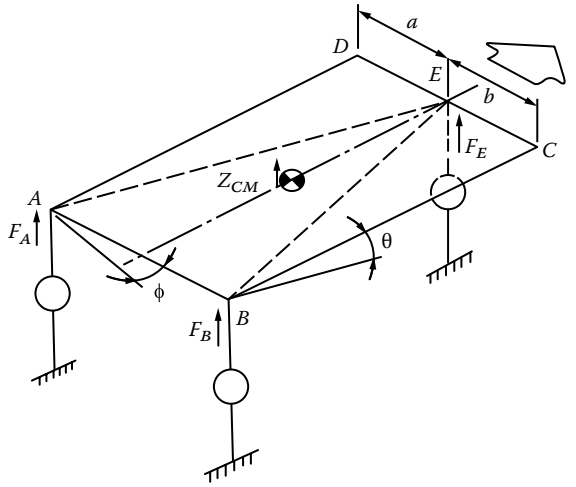


FIGURE 31.20 Equivalent "tricycle" configuration.

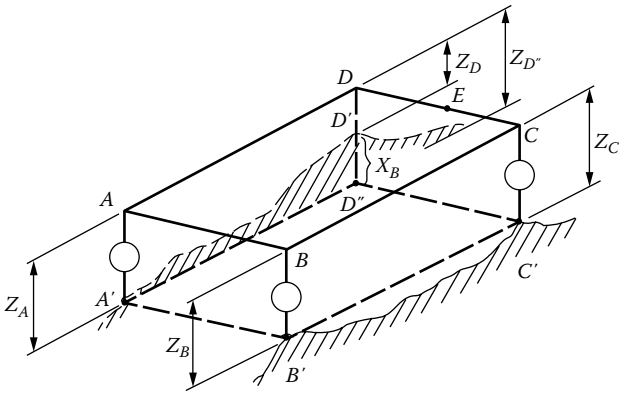


FIGURE 31.21 Introduction of additional state.

roll modes and four corner suspension deflections. Once the analytical solution to the full-car optimal suspension problem was found, it was possible to establish a number of related properties characterizing the optimal structure. For example, it was shown that all three optimal body modes have the highly desirable damping ratio of 0.7 that is achieved via "skyhook" damping. This holds for any set of PI_5 weights in Equation 31.16 and can be extended to even more general performance indices.

Under the following mild conditions, the original 3D problem can be decoupled into simpler pitch + heave and roll subproblems (see Figure 31.22)

$$a = b, \quad q_A = q_B, \quad q_C = q_D \tag{31.17}$$

where a and b are the lateral distances from the CM in Figure 31.19. It is also assumed that the ground roughness has been modeled according to a spatial road surface description from [84,85], which, as discussed in [39], conveniently facilitates the earlier decoupling. Moreover, if the previously established conditions (31.11) and (31.12) hold, then the pitch + heave dynamics can be further decoupled into front and rear quarter-car submodels as shown in Figure 31.22. In this way the original 3D problem is transformed into the much simpler, quarter-car models for which there is a wealth of results available, as discussed in previous sections. It should also be pointed out that, similar to the 2D case, it is now possible to consider 3D two-time scale and hierarchical control schemes, effects of flexible modes, preview, and different jerk-optimal PI augmentations.

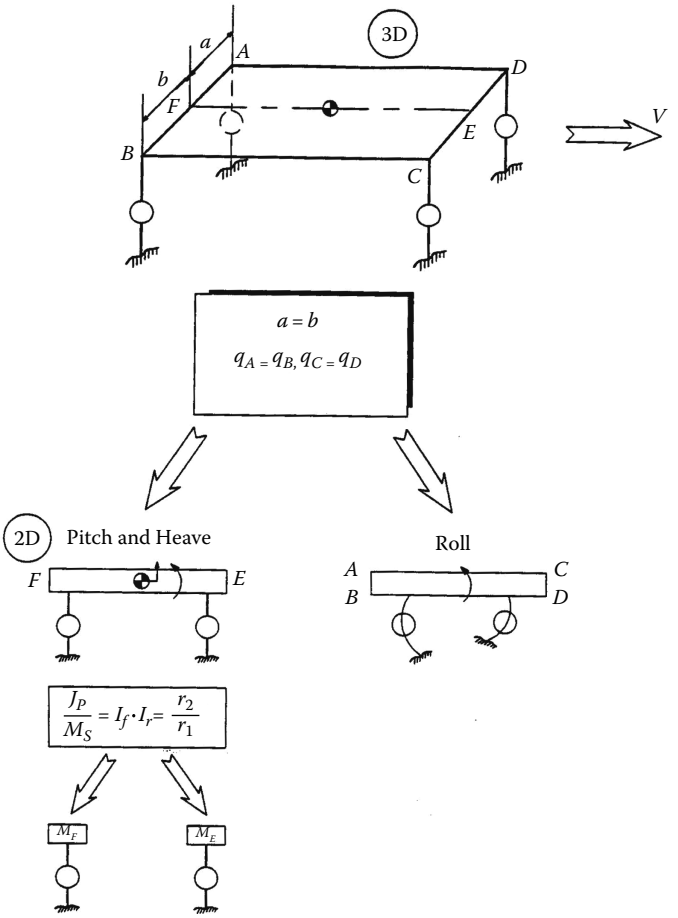


FIGURE 31.22 Front–rear–roll decomposition.

31.8 RELATED TOPICS

The preceding sections presented in a systematic fashion the fundamental characteristics of active suspensions. This was needed to establish key performance potentials and inherent limitations of different vehicle-suspension system models. A systematic way of presenting the results starting from the simplest possible and yet still relevant models is highly recommended since it leads to obtaining the fundamental insight, which can very easily be obscured by unnecessary details and case-to-case varying hardware specifics. The objective of this section is to briefly discuss developments in important areas that are directly or indirectly related to, or naturally follow after, these fundamental considerations.

31.8.1 Semi-Active Suspensions

The SA suspension concept was introduced in the early 1970s [41,86,87] in the form of variable, controllable damping. In this section we will somewhat loosely refer to "semi-active" as any essentially passive suspension concept, where a relatively small amount of external energy is used to improve the performance.

31.8.1.1 On–Off Semi-Active Damping

Since their introduction, SA dampers have been considered for diverse applications ranging from trains [88] to tractors and other off-road vehicles [89] to high-speed air-cushion vehicles [90,91] and military tanks [92]. Production applications of some very rudimentary SA concepts started in the early 1980s (see [24] of Chapter 22), with the introduction of the first variable damping shock absorbers. These variable dampers typically changed the damping from soft to firm and vice versa, through a manual or slow adaptive control and as such represented a very elementary form of the SA damping. The corresponding ride and handling performance improvements were marginal and sometimes even imperceptible, although some subsequent, more matured implementations show promise, especially in adapting to increased handling demands.

A somewhat more sophisticated and more demanding approach (in terms of implementation) is the so-called "on–off" SA strategy, which was first proposed in [90]. Simply stated, it switches the damper off, that is, creates a low or near-zero force, whenever the sprung and unsprung masses move in the same direction and the unsprung mass has a larger velocity. In any other situation the damper is set to "on" state, which produces large damper forces. The main idea of such a strategy was to reduce the sprung mass acceleration and motion, which was successfully demonstrated in early simulation and experimental studies [90,91,93,94]. However, care must be taken to prevent and smooth out too aggressive on–off switching, which may result in excessive jerks. This and similar refinements are a normal and necessary part of development from a relatively simple concept phase to actual production implementation.

During the recent years there has been considerable interest in the on–off SA concept in the industry. As a result, there appeared further improvements and refinements of the concept [95–98]. This is emerging as the first bona fide SA concept to see production [99,100].

31.8.1.2 Continuously Variable Semi-Active Damping

The original continuously variable SA modulation policy [41,86,87] is more challenging to implement than the earlier simple on–off concept, but at the same time it offers further performance improvement opportunities. It requires that the SA actuator continuously reproduces the (1-DOF LQ-optimal) skyhook damping force whenever this is possible in view of the passivity constraint. When this is not possible, the damper is simply turned off (here care should be exercised to avoid discontinuous-force switching that can lead to excessive jerk). The continuously variable SA policy was subsequently extended to 2-DOF, quarter-car, and more complex models, with simultaneous possible limiting of the damper maximum force leading to so-called "clipped" SA control [42,89,101–103].

While the clipped SA policy is in general not optimal [104,105], based on experience and prior results, it can be said that for most cases of practical significance, it is very close to optimal [102]. In its most basic form, the SA control problem can be cast in either linear state-space equation setup with nonlinear

constraints on SA damper forces or bilinear [42] with relatively simple inequality constraint on control variable—in this case SA actuator damping. It is also possible to represent the SA suspension design problem in the model predictive control (MPC)—hybrid control setting [105] with passivity-induced constraint as a natural switching parameter. Different (sub)optimal results are obtained depending on control horizon length and other MPC parameters. For control horizon of only one step in the future, the optimal control in this case coincides with the earlier clipped SA control [105].

Further improvements of the SA concept were also investigated through hardware modifications and extensions [106,107]. Usage of electrorheological (ER) and magnetorheological (MR) fluids has been proposed as a hardware alternative for the more conventional, variable-orifice hydraulic dampers [108,109]. From a practical standpoint, the MR concept appears especially promising for suspension and other vibration isolation applications (e.g., engine mounts), since it can operate on vehicle battery voltage, whereas the ER concept is based on high-voltage electric fields. There were also attempts to design the SA controllers using genetic algorithm [110], internal model control [111], dynamic programming [112], frequency-selective skyhook damping [113], neural networks [114], sliding mode controller [115], and preview control [116–120]. Actual production hardware and/or close-to-production hardware implementations are described in [99,100,113,121] and more recently in [122,123]. Additional related details and information can be found in Chapter 22 on advanced suspension hardware considerations.

31.8.1.3 Semi-active Spring

The concept of an incrementally SA spring, realized by means of on and off switching of additional gas spring accumulators, was introduced by [124]. This is a promising practical complement to SA damping, since, based on discussions in Section 31.5.2, both the suspension damping and stiffness should change to optimally adapt to different road/speed conditions. This can also be seen in Figure 31.23 reproduced from [125], which illustrates the performance of adaptive "passive" suspensions with the sprung mass natural frequency varying from 0.5 to 2 Hz and the corresponding damping ratio ζ ranging from 0 to 1. Clearly, the suspension stiffness has a significant influence on ride quality so that, similar to the fully active case, softer suspensions lead to more comfortable rides at the expense of increased suspension stroke requirements. Figure 31.23 also shows the corresponding optimal active boundaries. It can be seen that even such an idealized adaptive "passive" suspension cannot fully match the performance of the optimal active counterpart.

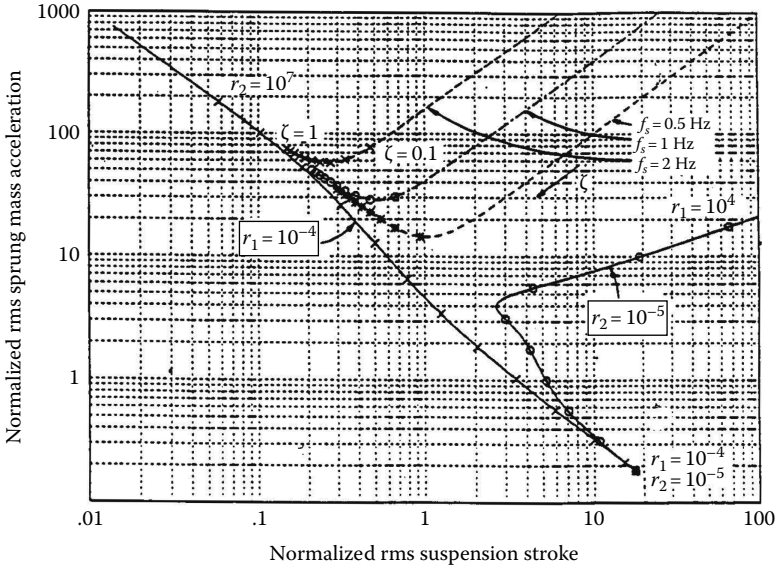


FIGURE 31.23 Comparison between quarter-car, 2-DOF models for passive suspension with f_s = [0.5, 1, 1.5] Hz, ζ = [0.1, 0.2, …, 1] and active suspension with $r_1 = 10^{-4}$ and $r_2 = 10^{-5}$.

31.8.2 State Estimation and Closed-Loop System Robustness

The underlying assumption in the preceding section has been that all the states are exactly known, which then gave us the best possible performance. In practice, some of the states may not be easily accessible. For the simple, quarter-car model, typical measurements may include sprung mass acceleration and suspension deflection, whereas the tire deflection is difficult to measure. Fortunately, the latter can often be neglected in the feedback control law without significant loss of performance, although there are operations when this cannot be done [24].

Based on the earlier measurements, it is possible to estimate the full-state vector using, for example, the Kalman filter in the context of the LQG methodology [126–129]. While such systems may reduce sensor requirements, it is important to remember that they typically lead to degradation in performance. For some realistic examples of systems using a single, rattlespace deflection measurement as in [127], it was shown [129] that such LQG policy can increase the LQ cost by more than 80%.

In addition, the LQG controller (with state observers or Kalman filters) is in general less robust with respect to modeling errors than the corresponding full-state counterpart. For a "typical" set of control gains, the gain margin of the suspension stroke-based LQG regulator can be only 0.2 dB, with the phase margin of 18° [129]. The corresponding LQ margins were ∞ dB and 100°. It should be pointed out that for a "softer" gain setting, even the full-state LQ regulator resulted in a small gain margin of only 1.5 dB and a phase margin of 33°. This is the consequence of a particular quarter-car structure, which included a parallel passive spring to (partially) support the vehicle. This additional spring leads to nonzero weightings on cross-terms between states and control in the PI, for which the "standard" LQ guarantees of ∝ gain margin and at least 60° of phase margin do not apply.

The preceding text indicates that it is important to consider closed-loop system robustness with respect to unmeasured states, structural changes (e.g., the introduction of a parallel supporting spring), as well as parametric uncertainties and, to a lesser degree, uncertainties due to sensor and actuator dynamics. To address various forms of uncertainties, a number of recently developed and computerized robust control techniques can be considered, with the ones based on frequency domain analysis being especially popular [130–132].

In particular, [132] investigated potential applications of H_∞ optimization in the context of a system-level active suspension design. It was found that, based on a performance mimicking Equation 31.10, the resulting controller introduced too much damping in the wheel-hop mode, since the H_∞ technique tends to minimize the peaks in the frequency-response curves. As might have been expected from the previous discussions, excessive damping will in general lead to deterioration of ride comfort performance. One of the potential benefits of H_∞-based approaches is for robustness analysis and enhancement. Especially promising here is the μ synthesis and analysis [133] that has already been applied to a number of automotive areas starting from, for example, [134], but it still has not been fully exploited in the context of active suspension design. Additional benefits are also possible by combining H_2 and H_∞ metrics [135]. For a full utilization of this mixed setting, one should leverage the H_∞ part for simultaneous optimization of a combined ride/handling/robustness trade-off.

31.8.3 Nonlinear Suspension Systems and Controls

Primary focus thus far was on optimal control methods applicable to linear vehicle models that are of course only approximately valid since the real vehicle system is necessarily nonlinear. For many operations, however, the linear system approximation is appropriate. This is also supported by experimental validations as discussed in [24]. However, there are situations that amplify the nonlinear effects. One example is created by discrete-event disturbances, such as single bumps or potholes, which can cause suspension bottoming—a highly nonlinear phenomenon. Another example is the always present dry friction, which may cause the suspension to lock on very smooth roads. The vehicle then essentially rides on tires. The present section gives a brief overview of different approaches to the nonlinear vehicle-suspension system design problem.

31.8.3.1 Nonlinear Control Design through Optimization

The nonlinear effects of suspension bottoming and the like were treated in [136] based on an approach that is similar in spirit to the optimal SA approach [102] discussed in 31.6.1. As was the case with [102], the method from [136] is computationally intensive, especially for higher-order vehicle models. However, with continuous improvements in computer processing power, this will be less of an issue in the future. Encouraged by favorable results with discrete disturbance responses, reference [136] argues for superiority of nonlinear over LQ-adaptive suspension algorithms. However, one could also present the reverse argument, especially if the full potentials of the adaptive strategies are taken into account [24].

Instead of competing alternatives, the two approaches can be viewed as complementing ingredients of an overall optimized suspension strategy. In this context, the LQ-optimal adaptive or gain-scheduling controls would be used for most driving situations with nonlinear policies being activated during excessive discrete-event disturbances and other special circumstances. An approach that would naturally blend the two is presented next.

31.8.3.2 Nonlinear Controls as an Extension of LQ Controls

From previous sections we know that the LQ setting offers a very attractive alternative and natural setting for solving the optimal or "best possible" active suspension design problem at the highest system level. This leads to fundamental results in terms of key limitations and potential benefits of such suspensions. However, one of the potential weaknesses of the LQ approach is that the actual hard constraints—such as rattlespace and tire deflection constraints—are approximated by their soft counterparts in the context of a quadratic penalty within an LQ PI. This leads to possibly more conservative results, especially when all suspension motion is confined to well within the allocated constraints.

To alleviate the above problem, Ref. [137] introduces additional quartic penalty terms for the earlier soft penalty constraints. This increases the penalty for excursions approaching the allowable space limits while at the same time reducing the same penalties for motions close to nominal, that is, close to static equilibrium. The resulting control law is summarized in Appendix 31.A. It has a desirable property that its linear term gains are identical to the well-known LQ gains ("downward compatibility"), thus recovering the LQ performance for relatively small motions, caused by relatively small road disturbances based on typical random road roughness characteristics. At the same time, the nonlinear control terms—in this case of a cubic variety—take care of excessive motions that can occur during encounters with sudden large bumps or potholes.

Illustrative examples in [137] show benefits of such a nonlinear suspension. In the case of a bump response, an effective "electronic tunable bumper" is providing additional flexibility in the tradeoffs between improved ride (lower acceleration levels) and allowable suspension and tire deflection. For random roads, there is a similar flexibility in terms of fundamental performance plots, which mimic Figure 31.9a and b in Section 31.5, but now for the nonlinear case [137]. This time the magnitude of road-induced roughness also alters these normalized plots, so that as the road disturbance increases the normalized rms tire deflection becomes smaller with simultaneous increase in the heave acceleration levels. This way, the nonlinear controller has some of the characteristics of its linear counterparts with gain scheduling. Of course, it is also possible to combine the earlier nonlinear controller with gain scheduling, which will inherit some additional desirable properties. The nonlinear portion will contain sudden large bumps and also offers some moderate adaptation to changes in prevailing road conditions. The gain-scheduling portion will then be used to accommodate or adapt to large changes in the prevailing road conditions.

31.8.3.3 Other Approaches to Nonlinear Suspension Control

There are many other nonlinear control-based approaches to active and SA suspension design. This includes sliding modes [138], adaptive controls [157], neural nets [26], feedback linearization [138],

and MPCs and hybrid controls [105,139], among others. In particular, reference [140] uses feedback linearization with backstepping for suboptimal active suspension design. It borrows some of the rationale from the earlier-mentioned decade-old optimization-based approaches, including a consideration of ride comfort versus rattlespace trade-offs, and adaptation to varying road roughness characteristics [141]. The centerpiece of this approach consists of an interesting albeit somewhat ad hoc nonlinear filter, which is less intuitive and straightforward to design than the optimization-based approaches. For example, [140] focuses on comfort versus rattlespace trade-offs, and it is not clear how easily one could incorporate the other important trade-offs such as the trade-off between handling (tire deflection) and ride comfort. Also, since there are no reported experimental results, it is not clear how robust this approach would be to parameter and system dynamics uncertainties and inaccuracies.

One of the most promising approaches to future advanced suspension design involves MPCs with their recent extensions in the form of hybrid controls [142–144]. The MPC control consists of using an appropriate plant model along with hard state, output and control constraints, which are in turn used to calculate optimal controls for an appropriate PI (typically in a quadratic or H_2 form) that spreads through a finite time horizon. Once the optimal control sequence has been found, only the first member of this sequence is actually implemented and the whole process of optimization then repeated. By implementing only the first step, one is in effect trying to contain unpredicted disturbances and model inaccuracies. As such, MPC is a very plausible and intuitively appealing approach that in a sense mimics control exerted by many living beings. At the same time, it is a natural extension of the earlier LQ optimization.

MPC controls have been successfully used for many years in the process industry, where update rates are relatively slow (of the order of seconds, minutes, or even hours), thus allowing for the earlier optimization to be performed on line. With ongoing progress in microprocessor infrastructure and MPC theory, it is now possible to consider other areas of application—such as aerospace and automotive—where control update rates are much faster (of the order of milliseconds). A prominent facilitator here was provided by development of the so-called explicit MPC methodology [145,146], where one performs all key optimizations off-line and stores the results in tables from where then actual controls are obtained by on line data entries based on plant states and various (measured disturbance and reference) inputs. The method was further extended to include nonlinear and hybrid systems, where the PI is now typically in the L_1 form and various nonlinearities and switching logic are represented by piecewise affine (PWA) functions [144].

Recent MPC/hybrid applications in the automotive industry include idle speed control [147], direct injection and stratified engine controls [148], transmission controls [149], traction control (TC) [150], and advanced active [139] and SA suspension control [105]. In particular, [139] is one of the first to effectively use and implement the MPC methodology in the context of active suspensions. The authors consider a quarter-car suspension model with road preview. The quadratic PI is similar to the earlier standard formulation (31.10) with an additional soft penalty on control force and a hard penalty on suspension travel or rattlespace. Also, a gain-scheduling type of road adaptation somewhat similar to [141] was used to facilitate easier satisfaction of the rattlespace hard constraint. The authors use more or less prevalent and most of the time preferable outer–inner loop approach based on the desired force generation and tracking (see Chapter 22, Section 22.3.3). The MPC optimization was performed on line in a laboratory half-car test rig, where the MPC (outer) loop was updated at a 10 ms rate. The resulting MPC-optimal force was then used as a reference signal for a sliding mode-based inner-loop controller implemented with an electrohydraulic setup operating at a 1 ms rate. The experimental results for smooth roads show similar benefits as for the standard LQ case from Section 31.5.2, whereas for harsh road profiles significant additional benefits resulted with MPC where rattlespace hard constraint prevents hitting the bump stops.

We conclude that, as a result of the substantial ongoing theoretical advances in the areas of nonlinear, model predictive, hybrid, and adaptive controls (cf., e.g., [144,151–155]), it is expected that the future will see increased applications of these and similar techniques in advanced suspension design.

Some early results in these directions have been mentioned earlier; additional attempts can be found in [156–159,141] and others.

31.8.4 Practical Considerations

The most direct way to actually implement the earlier optimal control-based suspension strategies is to use an inner-loop force control configuration. In this case, the outer/higher level provides the optimal desired or commanded force for the inner-loop, force-tracking control. This approach also applies in other cases, when the desired force can be generated via other control methods (e.g., classical and backstepping), which are not necessarily based on an explicit optimization. In practice, the active actuator should be supplemented by a more traditional passive counterpart consisting of spring and shock absorber or damper, typically geometrically in parallel with the active device. Of course, every practical active suspension implementation will also have an in-series compliance. In the case of the so-called narrow or low-bandwidth suspensions, this compliance is intentionally inserted to improve high-frequency isolation and reduce bandwidth requirements. (additional details can be seen in Chapter 22). They should work in tandem with the active device to deliver the desired force while at the same time minimizing the required energy expenditure (see Section 31.5.2.8). In addition, this approach leads to more robust and easier force-tracking implementation, since the presence of additional (passive) damping results in better preconditioning of the related system zeros that typically limit the inner-loop, force control performance (see Section 22.3.3). It also provides for a limp-home capability in case of active actuator failure.

For practical implementation, it is sometimes preferable to simplify the optimal strategies, thus leading to simpler software implementations at the expense of hopefully only slightly degraded performance. For example, in the case of quarter-car models, the full-state LQ-optimal regulator can be simplified to a three-term controller structure consisting of secondary suspension spring and damper (for body support and wheel-hop damping) and a skyhook-like damper for improved sprung mass damping and isolation. This is a suboptimal policy since it neglects tire deflection feedback. In many cases the effects of such an approximation are insignificant.

A similar approximation was used in quarter-car laboratory experiments with a wide-bandwidth electrohydraulic suspension leading to the results shown in Figure 31.24, where control gains have

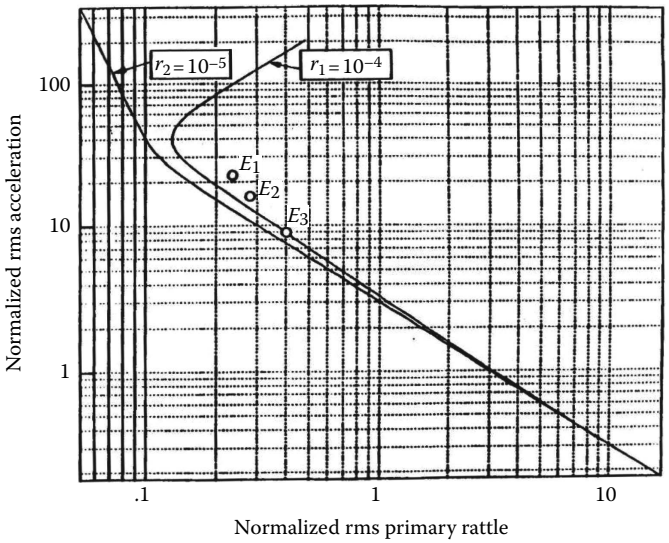


FIGURE 31.24 Optimal performance boundaries for $r_1 = 10^{-4}$ and $r_2 = 10^{-5}$ and experimental results E_1, E_2, E_3, for a quarter-car setup with an active suspension ($\rho = 12.5$, $\omega_{us} = 2\pi \cdot 12$ s^{-1}).

been adapted to the assumed varying road conditions ([24]). The comparison with superimposed optimal 2-DOF curves shows that, although not optimal, the experimental results follow the desired trends and are relatively close to the best possible performance. This gives credence to the underlying optimal analysis and proposed adaptive concepts.

More about practical aspects of advanced suspension design and implementation can be found in Chapter 22. This chapter contains description and comparison of different approaches based on force and displacement inner loops. The chapter also discusses related hardware concepts and actuator issues as well as challenges associated with different practical realizations. Additional information can be found in [160].

31.8.5 Other Aspects

Traditionally, the main focus of active and SA suspension control was on applications dealing with models limited to rigid body dynamics only. This includes vehicle heave, pitch, and roll modes. A consideration of some flexible mode effects can be found in [68,161].

It is believed that the earlier optimal control and robustness results provide a solid base for fault-tolerant control design, which typically constitutes the next important step in developing a product. This includes the generation of suitable dynamic models needed for system diagnostics and failure accommodation strategies.

While all these aspects are extremely important and must be addressed prior to production, they are beyond the scope of this chapter. The interested reader is referred to ongoing research publications in this dynamic field.

In addition to fault analysis and accommodation, the areas of fuzzy logic and neural network control applications can also benefit from the earlier optimal control findings. This includes the choice of proper input variables and overall control policy that, at least to a first degree, may mimic the LQ-optimal logic. Due to their practical and intuitive appeal, these techniques, and especially fuzzy logic, may provide an efficient mechanism to bring the more structured control approaches "home"—in the production car suspensions, power trains, brakes, and other devices.

The present emphasis was on optimal control applications to the design of advanced automotive suspensions, although many conclusions apply in general to other transportation areas. For more specific overviews of results pertinent to other areas, consult the following papers and references therein: commercial vehicle applications [162,163], rail vehicle dynamics [164–166], magnetically suspended vehicles [164,167,168], off-road and military vehicles [169], race cars [170], surface-effect ships [171], and other vehicles.

It should be mentioned that active suspensions can be integrated with other advanced vehicle subsystems into more effective, overall integrated and interactive vehicle control systems, which can significantly improve many vehicle functions especially with respect to vehicle stability in safety-critical emergency maneuvers.. References [172–176] combine active and SA suspensions with four-wheel steer (4WS), antilock braking (ABS), TC, and four-wheel drive (4WD) to achieve superior performance, handling control, ride comfort, and overall vehicle safety during acceleration/braking on straight-line driving as well as, in turns, on dry or slippery roads.

An example of using SA damping to improve vehicle handling and control on a rough slippery road is provided in [177]. Here both decentralized and coordinated controls between the SA damping and dynamic stability control have been investigated. It was found that the coordinated controls offer additional significant benefits in this case in terms of vehicle handling and safety. Present trend toward increased integration and interaction will continue to accelerate. This in turn will lead to increased prospects for future production introductions of active and SA suspensions. Their increased usage will facilitate superior and new functions that will impact many aspects of vehicle performance, comfort, and safety.

ACKNOWLEDGMENTS

Help from my Ford Research colleagues Dr. Jahan Asgari, Mr. Mike Fodor, and especially Dr. H. Eric Tseng is gratefully acknowledged. In addition, Dr. Niklas Karlsson contributed with an appendix on relevant nonlinear suboptimal controls.

31.A APPENDIX: NONLINEAR SUBOPTIMAL CONTROL

Consider the cost function

$$J = \frac{1}{2}\int_0^\infty \left(x^T Q x + u^T R u + \varepsilon f(x) \right) dt \tag{31.A.1}$$

where $\varepsilon f(x)$is a fourth-order polynomial in the state x, subject to

$$\dot{x} = Ax + Bu + \Gamma w \tag{31.A.2}$$

The solution to (31.A.1) and (31.A.2) was first derived in [137]; a more thorough presentation together with some important stability results is available in [178].

Here follows a step-by-step scheme for computing a suboptimal controller $u_{subopt}(x)$ to (31.A.1) and (31.A.2). The optimal controller can be expressed as

$$u_{opt} = -R^{-1}B^T \left(Px + \sum_{k=1}^{\infty} \varepsilon^k g^{(k)}(x) \right)$$

The suboptimal controller, on the other hand, is defined as

$$u_{subopt} = -R^{-1}B^T \left(Px + \varepsilon g^{(1)}(x) \right)$$

To compute u_{subopt}, the matrix P is first obtained as the stabilizing solution of the following Riccati equation:

$$PA + A^T P - PBR^{-1}B^T P + Q = 0$$

It can then be shown that $g^{(1)}$ is the solution to the linear equation

$$\frac{\partial g^{(1)}}{\partial x}(A - BR^{-1}B^T P)x + (A^T - PBR^{-1}B^T)g^{(1)}(x) = -\frac{1}{2}\left(\frac{\partial f}{\partial x} \right)^T \tag{31.A.3}$$

with the boundary condition $\lim_{t\to\infty} g^{(1)}(x,t) = 0$. Since $f(x)$ is a fourth-order polynomial in the state x, it is clear by inspection of Equation 3 that $g^{(1)}$ is a vector of polynomials of order three. Assume the state has dimension four. The function $g^{(1)}$ can then be written as

$$g^{(1)}(x) = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} \begin{bmatrix} G^1_{ijk} x_i x_j x_k \\ G^2_{ijk} x_i x_j x_k \\ G^3_{ijk} x_i x_j x_k \\ G^4_{ijk} x_i x_j x_k \end{bmatrix}$$

Similarly, $f(x)$ can be written as the sum

$$f(x) = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4}\sum_{l=1}^{4} F_{ijkl}x_i x_j x_k x_l$$

Now, define the vectors ξ and η from the coefficients of $f(x)$ and $g^{(1)}$ as follows:

$$\xi = \left[G^1_{111}, G^2_{111}, G^3_{111}, G^4_{111}, G^1_{112}, \ldots, G^4_{112}, G^1_{114}, \ldots, G^1_{444}, \ldots, G^4_{444} \right] \in \Re^{256}$$

and

$$\eta = 2 \cdot \left[F_{1111}, \ldots, F_{1114}, F_{1121}, \ldots, F_{1124}, F_{1131}, \ldots, F_{4441}, \ldots, F_{4444} \right] \in \Re^{256}$$

The sought for vector ξ is obtained as the solution to

$$\hat{S}\xi = \eta,$$

where $\hat{S} = S^T \oplus S^T \oplus S^T \oplus S^T \in \Re^{256\times 256}$ and $S = A - BB^T P$. The symbol $\oplus$ denotes the Kronecker sum operator, defined by $S^T \oplus S^T = (I_4 \otimes S^T) + (S^T \otimes I_4)$.

A more general case is discussed in [178], where it is also shown that the suboptimal controller u_{subopt} under certain conditions guarantees a stable closed-loop system.

Example

Consider, for example, the active control of the quarter car described in Section 31.5.2. The dynamics is described by the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -(20\pi)^2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 10 \\ 0 \\ -1 \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Furthermore, let the cost function be

$$J = \frac{1}{2}\int_0^\infty \left(x^T Q x + u^2 + \varepsilon f(x) \right) dt \quad \text{where } Q = \begin{bmatrix} 1100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$f(x) = 12{,}375{,}000 \cdot x_1^4 + 1{,}125{,}000 \cdot x_3^4$ and $\varepsilon = 0.25$. Note that the quartic terms penalize x_1 (the tire deflection) and x_3 (the suspension deflection) to the fourth order. This choice of $f(x)$ corresponds to a vector η where the only nonzero elements are $F_{1111} = 12{,}375{,}000$ and $F_{3333} = 1{,}125{,}000$.

The suboptimal controller is obtained as

$$u = ax_1 + bx_2 + cx_3 + dx_4 + \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} D_{ijk}x_i x_j x_k,$$

where $a = 6.0842$, $b = -0.5478$, $c = 10.0$, $d = 4.4375$ and where the coefficients D_{ijk} are given in the following table:

ijk	D_{ijk}	*ijk*	D_{ijk}	*ijk*	D_{ijk}	*ijk*	D_{ijk}
111	$1.16 \cdot 10^{4}$	123	$-4.80 \cdot 10^{2}$	222	$-1.43 \cdot 10^{-1}$	244	-3.35
112	$-7.95 \cdot 10^{2}$	124	$-2.86 \cdot 10^{1}$	233	2.78	333	$1.41 \cdot 10^{4}$
113	$4.06 \cdot 10^{4}$	133	$3.57 \cdot 10^{4}$	224	$2.74 \cdot 10^{-1}$	334	$6.07 \cdot 10^{3}$
114	$5.70 \cdot 10^{3}$	134	$1.02 \cdot 10^{4}$	233	$-2.75 \cdot 10^{2}$	344	$1.19 \cdot 10^{3}$
122	$-3.34 \cdot 10^{-1}$	144	$9.84 \cdot 10^{2}$	234	$-4.53 \cdot 10^{1}$	444	$9.03 \cdot 10^{1}$

REFERENCES

1. Morman, Jr. K. N. and F. Giannopoulas, Recent advances in the analytical and computational aspects of modeling active and passive vehicle suspensions. In Kamal, M. M. and Wolf Jr., J. A. (eds.), *Computational Methods in Ground Transportation Vehicles*. Vol. AMD 50, ASME, New York, 1982, pp. 75–115.
2. Bender, F. K., D. C. Karnopp, and I. L. Paul, On the optimization of vehicle suspensions using random process theory. ASME Paper No. 67-Tran-12, 1967.
3. Bender, F. K., Optimum linear control of random vibrations. *Proceedings of the Joint American Control Conference*, Philadelphia, USA, June, 1967, pp. 135–143.
4. Bender, E. K., Optimization of the random vibration characteristics of vehicle suspensions using random process theory. ScD. Thesis, MIT, Cambridge, MA, 1967.
5. Karnopp. D. C. and A. K. Trikha, Comparative study of optimization techniques for shock and vibration isolation. *ASME Journal of Engineering for Industry,* 91(4), 1128–1132, 1969.
6. Williams, D. A. and P. G. Wright, Vehicle suspension system. U.S. Patent 4 625 993, 1986.
7. Thompson, A. G., Design of active suspensions. *Proceedings of the Institution of Mechanical Engineering*, 185(36), 553–563, 1970–1971.
8. Karnopp, D. C., Two contrasting versions of the optimal active vehicle suspension. *ASME Monograph*, DSC-1, 341–346, 1985.
9. Williams, R. A., Active suspensions; classical or optimal? *Proceedings of the 9th IAVSD Symposium*, Linkoping, Sweden, 1985, pp. 607–620.
10. Michelberger, P., J. Bokor, A. Keresztes, and P. Varlaki, Design of active suspension system for road vehicles; an eigenstructure assignment approach. SAE Paper No. 905146, 1990.
11. Smith, M. C., Achievable dynamic response for automotive active suspensions. *Vehicle System Dynamics*, 24, 1–33, 1995.
12. Orlandea, N. and M. A. Chace, Simulation of a vehicle suspension with the ADAMS computer program. SAE Paper No. 770053, 1977.
13. Grubel, G. and W. Kortum, Towards a coherent technology for computational vehicle system dynamics. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Yokohama, Japan, 1992, pp. 48–55.
14. Schiehlen, W., Multibody dynamics software for controlled vehicle simulation. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Yokohama, Japan, 1992, pp. 37–42.
15. Kortum, W., W. Rulka, and W. Schwartz, Analysis and design of controlled vehicles using multibody simulation models. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 85–92.
16. Lamps. M. F. and E. C. Ekert, Improving the suspension design process by integrating multibody system analysis and design of experiments. SAE Paper No. 930264, 1993.
17. Fritz, C., M. Hassenforder, D. Lienhardt, and G. Gissinger, A vehicle suspension study with a versatile software package called PROUESSE. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 93–98.
18. Foag, W. and G. Grubel, Multi-criteria control design for preview vehicle-suspension systems. *Proceedings of the 10th IFAC World Congress*, Munich, Germany, 3, pp. 190–196, 1987.
19. Zaremba, A., R. Hampo, and D. Hrovat, Control synthesis for an automotive active suspension using constrained optimization. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994.
20. Mastinu, G., Automotive suspension design by multi-objective programming. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994.

21. Hedrick, J. K., C. F. Billington, and D. A. Dreesbach., Analysis, design and optimization of high speed vehicle suspensions using state variable techniques. *ASME Journal of Dynamic Systems, Measurement, and Control*, 96(2), 193–203, June 1974.
22. Guenther, D. R. and C. T. Leondes, Synthesis of a high-speed tracked vehicle suspension system. Part I: Problem statement, suspension structure, and decomposition; and, Part II: Definition and solution of the control problem. *IEEE Transactions on Automatic Control*, AC-22(2), 158–172, April 1977.
23. Lu, X. P., H. L. Li, and P. Papalambros, A design procedure for the optimization of vehicle suspensions. *International Journal of Vehicle Design*, 5(1/2), 129–142, 1984.
24. Hrovat, D., Survey of advanced suspension developments and related optimal control applications. *Automatica*, 33(10), 1781–1817, 1997.
25. Hampo, R. J. and K. A. Marko, Investigation of the application of neural networks to fault tolerant control of an automotive suspension system. *Proceedings of the American Control Conference*, Chicago, IL, 1992, pp. 11–15.
26. Feldkamp, L. A., G. V. Puskorius, L. I. Davis. Jr, and K. Yuan, Neural control systems trained by dynamic gradient methods for automotive application. *Proceedings of the International Joint Conference on Neural Networks*, Baltimore, MD, Vol. II, 1992, pp. 798–804.
27. Nagai. M. and A. Moran, Optimal rear suspension preview control of nonlinear vehicles using neural networks. *Proceedings of the 12th IFAC World Congress*, Sydney, Australia, 1993.
28. Presser, S., A. Vikas, A. Woehler, and H. P. Willumeit, Identification of fuzzy-control systems by evolution strategy. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 116–121.
29. Marsh, C., T. J. Gordon, and Q. H. Wu., Application of learning automata to controller design in slow-active automobile suspensions. *Vehicle System Dynamics*, 24, 597–616, 1995.
30. Kashani, R., J. E. Strelow, M. Eiler, and M. Aryakula, LQG-based fuzzy logic control of active suspension. *ASME Journal of Dynamic Systems, Measurement and Control*, 58, pp. 69–76, 1996.
31. Smith, C. C., D. Y. McGehee, and A. J. Healey, The prediction of passenger riding comfort from acceleration data. *ASME Journal of Dynamic Systems, Measurement, and Control*, 100, 34–41, March 1978.
32. Hedrick, J. K., R. K. Ravera, and J. R. Amderes, The effect of elevated guideway construction tolerances on vehicle ride quality. *ASME Journal of Dynamic Systems, Measurement, and Control*, 97(4), 408–416, December 1975.
33. Fearnsides, J. J., J. K. Hedrick, and H. Firouztash., Specification of ride quality criteria for transportation systems: The state of the art and a new approach. *High Speed Ground Transportation Journal*, 8(2), 125–132, 1974.
34. Hoberock, L. L., A survey of longitudinal acceleration comfort studies in ground transportation vehicles. *ASME Journal of Dynamic Systems, Measurement, and Control*, 99(2), 76–84, June 1977.
35. Anonymous, A guide to the evaluation of human exposure to whole body vibration. ISO/DIS 2631, International Standard Organization, New York, 1972.
36. Smith, C. C., On using the ISO standard to evaluate the ride quality of broad-band vibration spectra in transportation vehicles. *ASME Journal of Dynamic Systems, Measurement, and Control*, 98(4), 440–443, December 1976.
37. Klinger, D. L., N. K. Cooperrider, J. K. Hedrick, R. C. White, A. Cilzado, M. Sayers, and D. Wormley, Guideway vehicle cost reduction, parts I and II. National Technical Information Service Reports DOT-TST-75–95, 1976.
38. Stephens, D. G., Comparative vibration environments of transportation vehicles. In Berman, A. and Hannibal, A. J. (eds.), *Passenger Vibration in Transportation Vehicles*, ASME Publication AMD, Chicago, IL, Vol. 24, September 1977, pp. 59–72.
39. Hrovat, D., Applications of optimal control to advanced automotive suspension design. *ASME Journal of Dynamic Systems, Measurement, and Control*, (Special Issue commemorating 50 years to the DSC division), 115, 328–342, 1993.
40. Asgari, J. and Hrovat, D., Bond graph models of vehicle 2D ride and handling dynamics. In Velinsky, S. A., R. H. Fries, I. Hague, and D. Wang. (eds.), *Advanced Automotive Technologies*, ASME Publication, DE Vol. 40, December 1991, pp. 391–406.
41. Crosby, M. and Karnopp, D. C., The active damper—A new concept for shock and vibration control. In *The Shock and Vibration Bulletin,* Part H, Washington, DC, 43(4), pp. 119–133, 1973.
42. Hrovat, D., Optimal passive vehicle suspensions. PhD thesis, University of California, Davis, CA, May 1979.
43. Hrovat, D., D. L. Margolis, and M. Hubbard., Suboptimal semi-active vehicle suspensions. *Proceedings of the 1980 JA*CC, San Francisco, CA, August 1980.

44. Anderson, B. D. O. and S. Vongpanitlers., *Network Analysis and Synthesis*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1973.
45. Hrovat, D., A class of active LQG optimal actuators. *Automatica*, 18(1), 117–119, January 1982.
46. Sevin, E. and W. D. Pilkey, *Optimum Shock and Vibration Isolation*, The Shock and Vibration Information Center, United States Department of Defense, Washington, DC, 1971.
47. Hrovat, D. and D. L. Margolis., Realistic road-truck systems simulation using digital computers. *Proceedings of the Winter Computer Simulation Conference*, Sacramento, CA, 1975.
48. Healey, A. J., Digital processing of measured vibration data for automobile ride evaluation. In Berman, A. and Hannibal, A. J. (eds.), *Passenger Vibration in Transportation Vehicles*, ASME Publication AMD, Vol. 24, September 1977, pp. 1–17.
49. Hrovat, D. and M. Hubbard, A comparison between jerk optimal and acceleration optimal vibration isolation. *Journal of Sound and Vibration*, 12(2), 201–210, July 1987.
50. Hrovat, D., and M. Hubbard, Optimum vehicle suspension minimizing RMS rattlespace, sprung-mass acceleration and jerk. *ASME Journal of Dynamic Systems, Measurement, and Control*, 103(3), 228–236, September 1981.
51. Davison, E. J. and P. Patel, Application of the robust servo mechanism controller to systems with periodic tracking/disturbance signals. *International Journal of Control*, 47(1), 111–127, January 1988.
52. Boyd, S., L. El Ghaoui., E. Feron., and V. Balakishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, ISBN 0–89871–334-x, Philadelphia, PA, 1994.
53. Fialho, I. and G. J. Balas, Road adaptive active suspension design using linear parameter-varying gain-scheduling. *IEEE Transactions on Control Systems Technology*, 10(1), 43–54, January 2002.
54. Chen, H. and K. Guo, An LMI approach to multiobjective RMS gain control for active suspensions. *Proceedings of the 2001 American Control Conference*, Arlington, VA, 2001, pp. 2646–2651.
55. Becker, G., A. Packard, Robust performance of linear parametrically varying systems using parametrically-dependent linear feedback. *Systems and Control Letters*, 23(3), 205–215, September 1994.
56. Tran, M. and D. Hrovat, Application of gain-scheduling to design of active suspensions. *Proceedings of the 1993 IEEE Conference on Decision and Control*, San Antonio, TX, 1993, pp. 1030–1035.
57. Hrovat, D., Influence of unsprung weight on vehicle ride quality. *Journal of Sound and Vibration,* 124, 497–516, 1988.
58. Hedrick, J. K. and T. Butsuen, Invariant properties of automotive suspensions. *Proceedings of the Institution of Mechanical Engineering*, 204(1), 21–27, 1990.
59. Hrovat, D., Optimal active suspension structures for quarter-car vehicle models. *Automatica*, 26(5), 845–860, 1990.
60. Den Hartog, J. P., *Mechanical Vibrations*. McGraw-Hill, New York, 1956.
61. Hrovat, D., P. Barak, and M. Rabins, Semi-active versus passive or active tuned mass dampers for structural control of buildings. *ASCE Journal of Engineering Mechanics*, 109(3), 691–705, 1983.
62. Moran, T., AUTOS ON MONDAY/Technology; A new suspension's magnetic appeal, New York Times, Section D, p. 10, October 11, 2004.
63. Parison, J. A. and L. D. Knox., Wheel damping. Bose Corporation, US Patent 6,364,078, April 2002.
64. Salman, M. A., A. Y. Lec, and N. M. Boustany, Reduced-order design of active suspension control. *ASME Journal of Dynamic Systems, Measurement, and Control*, 112, 604–610, 1990.
65. Krtolica, I. T., U. Ozguner, H. Chan, and D. Hrovat, Two time-scale analysis of active suspension control of 2D/4DOF half-car model. Final Report, Ohio State University, Columbus, OH, 1990.
66. Fathy, H. K., A. G. Ulsoy, P. Y. Papalambros, and D. Hrovat, Nested plant/controller optimization with application to combined passive/active automotive suspensions. *Proceedings of the 2003 American Control Conference*, Denver, CO, 4, pp. 3375–3380, June 2003.
67. Thompson, A. G., An optimal suspension for an automobile on a random road. SAE Paper No. 790478, 1979.
68. Hac, A., Stochastic optimal control of vehicles with elastic body and active suspension. *ASME Journal of Dynamic Systems, Measurement, and Control*, 108, 106–110, June 1986.
69. Krtolica, R. and D. Hrovat, Optimal active suspension control based on a half-car model: An analytical solution. *IEEE Transactions on Automatic Control*, AC-37, 528–532, 1992.
70. Tomizuka, M., Optimum linear preview control with application to vehicle suspension—Revisited. *ASME Journal of Dynamic Systems, Measurement, and Control*, 98(3), 309–315, September 1976.
71. Hrovat, D., Optimal suspension performance for 2D vehicle models. *Journal of Sound and Vibration*, 145(I), 93–110, 1991.
72. Rutledge, D. C., Models for jerk-optimal vehicle suspensions. PhD thesis, University of California, Davis, CA, 1990.

73. Ono. E., M. Yamashita, and Y. Yamauchi, Distributed hierarchy control of active suspension system. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 373–378.
74. Hac, A., Optimal linear preview control of active vehicle suspension. *Vehicle System Dynamics*, 21, 167–195, 1992.
75. Araki, Y., M. Oya, and H. Harada, Preview control of active suspension using disturbance of front wheel. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 299–304.
76. Pilbeam, C. and R. S. Sharp, Performance potential and power consumption of slow-active suspension systems with preview. *Vehicle System Dynamics*, 25, 169–183, 1996.
77. Sharp, R. S. and C. Pilbeam, On the ride comfort benefits available from road preview with slow-active car suspensions. *Vehicle System Dynamics*, 23 (*Supplement on Proceedings of the 13th IAVSD Symposium*), 437–448, 1994.
78. Huisman, R. G. M., F. E. Veldpaus, J. G. A. M. van Heck, and J. J. Kok., Application of a preview controlled active suspension to a (non)linear 2-D truck model. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 293–298.
79. Hac, A., Decentralized control of active vehicle suspensions with preview. *ASME Journal of Dynamic Systems, Measurement, and Control*, 117, 478–483, 1995.
80. Barak, P. and D. Hrovat, Application of the LQG approach to design of an automotive suspension for 3D vehicle models. *Proceedings of the International Conference on Advanced Suspensions, IMECHE*, London, Great Britain, 1988.
81. Barak, P., On a ride control algorithm for heave, pitch and roll motions of a motor vehicle. PhD Thesis, Department of Mechanical Engineering, Wayne State University, Detroit, MI, 1985.
82. Chalasani, R. M., Ride performance potential of active suspension systems-Part II: Comprehensive analysis based on a full-car model. *ASME Monograph,* AMD-80, DSC-1, 1986, pp. 619–626.
83. Hrovat, D., Optimal active suspensions for 3D vehicle models. *Proceedings of the 1991 American Control Conference*, Boston, MA, 1991, pp. 1534–1541.
84. Parkhilovskii, I. G., Investigations of the probability characteristics of the surfaces of distributed types of roads. *Autumn Promotion*, 8, 18–22, 1968.
85. Rill, G., The innuence of correlated random road excitation processes on vehicle vibration. *Proceedings of the 8th IAVSD Symposium on the Dynamics of Vehicles on Roads and on Railway Tracks*, Cambridge, MA, August 1983, pp. 449–459.
86. Karnopp, D. C. and M. I. Crosby, System for controlling the transmission of energy between spaced members. U.S. Patent No. 3 807 678, 1974.
87. Karnopp, D. C., M. J. Crosby, and R. A. Harwood, Vibration control using semi-active force generators. *ASME Journal of Engineering Industry*, 96(2), 1974.
88. Klinger, D. L., N. K. Cooperrider, J. K. Hedrick, R. C. White, A. Cilzado, M. Sayers, and D. Wormley, Guideway vehicle cost reduction, parts I and II. National Technical Information Service Reports DOT-TST-75–95, 1976.
89. Roley, D. G., Tractor cab suspension performance modeling. PhD Thesis, Department of Agricultural Engineering, University of California, Davis, CA, 1975.
90. Margolis, D. L., J. L. Tylee, and D. Hrovat., Heave mode dynamics of a tracked air cushion vehicle with semiactive airbag secondary suspension. *ASME Journal of Dynamic Systems, Measurement, and Control*, 97(4), 399–407, 1975.
91. Margolis, D. L. and D. Hrovat, Semi-active heave and pitch control of a high speed tracked air cushion vehicle. *Intersociety Transportation Conference*, Los Angeles, CA, 1976.
92. Miller. L. R. and C. M. Nobles, The design and development of a semi-active suspension for military tank. SAE Paper No. 881133, 1988.
93. Hrovat, D. and D. L. Margolis, An experimental comparison between semi-active and passive suspensions for air-cushion vehicles. *International Journal Vehicle Design*, 2(3), 308–321, 1981.
94. Krasnicki, E. J., The experimental performance of an ON-OFF active damper. *Shock and Vibration Bulletin*, 51, 125–131, 1981.
95. Miller, L. R., The effect of hardware limitations on an on/off semi-active suspension. *Proceedings of the International Mechanical Engineering. International Conference on Advanced Suspensions*, Paper No. C442/88, London, pp. 199–206, 1988.
96. Ivers, D. E. and L. R. Miller, Experimental comparison of passive, semi-active on/off, and semi-active continuous suspensions. SAE Paper No. 892484, 1989.

97. Crawford, I. L., Semi-active suspension system. *Proceedings of the Conference on Advanced Vehicle Systems*, Strasbourgh, France, 1988, pp. 602–619.
98. Decker, H. and W. Schramm, An optimized approach to suspension control. SAE Paper No. 900 661, 1990.
99. Emura, J., S. Kakizaki, F. Yamaoka, and M. Nakamura., Development of the semi-active suspension system based on the sky-hook damper theory. SAE Paper No. 940863, 1994.
100. Higashiyama, K., T. Hirai, S. Kakizaki, and M. Hiramoto., Development of the active damper suspension. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 331–336.
101. Siebenhaar, A., The semi-active regulator. PhD Thesis, University of California, Davis, CA, 1975.
102. Hrovat, D., D. L. Margolis, and M. Hubbard, An approach toward the optimal semi-active suspension. *ASME Journal Dynamic Systems*, *Measurement and Control*, 110(3), 288–296, 1988.
103. Butsuen, T. and J. K. Hedrick, Optimal semi-active suspensions for automotive vehicles: The 1/4 car model. *Advanced Automotive Technology*, DSC-3, 305–319, 1989.
104. Tseng, H. E. and J. Karl Hedrick, Semi-active control laws—Optimal and sub-optimal. *Vehicle System Dynamics*, 23(7), 545–569, October 1994.
105. Giorgetti, N., A. Bemporad, H. E. Tseng, and D. Hrovat, Hybrid model predictive control application towards optimal semi-active suspension. *International Symposium on Industrial Electronics* (ISIE 2005), Mini Track on Automotive Control, Dubrovnik, Croatia, June 2005.
106. Ivers, D. E. and L. R. Miller, Semi-active suspension technology: An evolutionary view. *Advanced Automotive Technologies*, presented at the Winter Annual Meeting, Atlanta, GA, DE-Vol. 40, 1991, pp. 327–346.
107. Wallentowitz, H. and D. Konik., Actively influenced suspension systems. *Proceedings of the 3rd International Conference on Vehicle Dynamics and Powertrain Engineering*, Strasbourg, France, 1991.
108. Sturk, M., X. M. Wu, and J. Y. Wong, Development and evaluation of a high voltage supply unit for electrorheological fluid dampers. *Vehicle System Dynamics*, 24, 101–121, 1995.
109. Pinkos, A., E. Shtarkman, and T. Fitzgerald, An actively damped passenger car suspension system with low voltage electro-rheological magnetic fluid. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 311–317.
110. Yeh, E. C., S. H. Lu, and C. C. Chen, A genetic algorithm based fuzzy system for semi-active suspension system design. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 189–194.
111. Park. S. and J. Koo, Design of IMC structured automotive semi-active controller. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 183–188.
112. Muijderman. J. H. E. A., F. E. Veldpaus, and J. J. Kok, A semi-active suspension system based on dynamic programming. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 177–182.
113. Konik, D., W. Bauer, K. J. Huber, B. Jordan, S. Kolbel, J. Scharf, S. Schopp, and M. Wimmer, Electronic damping control with continuously working damping valves (EDCC)—System description and functional improvements. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Aachen, Germany, 1996, pp. 87–104.
114. Moran. A., T. Hasegawa, and M. Nagai, Continuously controlled semi-active suspension using neural networks. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 305–310.
115. Isobe, O., T. Kawabe, Y. Watanabe, Y. Miyasato, and S. Hanba., Sliding mode controller for semi-active suspension system for commercial vehicles. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Aachen, Germany, 1996, pp. 37–46.
116. Kawagoe, K., *Semi-Active Control and Optimum Preview Control with Application to Vehicle Suspensions*. Research Publication, Iguchi Laboratory, Tokyo University, Tokyo, Japan, 1983.
117. Kawagoe, K. and M. Iguchi, Semi-active control and optimum preview control applications to vehicle suspension. *JSAE Review*, 17, 24–31, 1985.
118. Hac, A. and I. Youn, Optimal semi-active suspension with preview based on a quarter car model. *Vibration, Acoustics, Stress, and Reliability in Design*, 114(1), 84–92, January 1992.
119. Jezequel, L. and V. Roberti, Study of a nonlinear preview semi-active suspension system. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Yokohama, Japan, 1992, pp. 123–129.
120. Huisman, R. G. M., F. E. Veldpaus, J. G. A. M. van Heck, and J. J. Kok, Preview estimation and control for (semi-) active suspensions. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Yokohama, Japan, 1992.

121. Irmscher, S., E. Hees, and T. Kutsche, A controlled suspension system with continuously adjustable damping force. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 325–330.
122. Alexandridis, A. A. and J. P. Goldasz, *7th Conference on Active Noise and Vibration Control Methods*, Wigry, Poland, June 8–12, 2005.
123. Modie, M. O. and A. Hac, Closed-loop yaw control of vehicles using magneto-rheological dampers. *SAE SP-1537*, 2000, SAE Paper 2000–01–0137.
124. Hubbard, M. and D. Margolis, The semi-active spring, is it a viable concept? *Intersociety Conference on Transportation*, Los Angeles, CA, 1976.
125. Tseng, T. and D. Hrovat, Some additional characteristics of optimal active suspensions for quarter-car models. Ford Motor Company Research Report SR.89–95, Dearborn, MI, 1989.
126. Alexandridis, A. A. and T. R. Weber, Active vibration isolation of truck cabs. *Proceedings of the American Control Conference*, San Diego, CA, 1984, pp. 1199–1208.
127. Yue, C., T. Butsuen, and K. Hedrick, Alternative control laws for automotive active suspensions. *Proceedings of the American Control Conference*, Atlanta, USA, June 1988, pp. 2373–2376.
128. Ulsoy, A. G. and D. Hrovat, Stability robustness of LQG active suspensions. *Proceedings of the 1990 American Control Conference*, San Diego, CA, 1990, pp. 1347–1356.
129. Ulsoy, A. G., D. Hrovat, and T. Tseng, Stability robustness of LQ and LQG active suspensions. *ASME Journal of Dynamic Systems, Measurement and Control*, 116(l), 123–131, 1994.
130. Kiriczi, S. and R. Kashani, Control of active suspension with parameter uncertainty and non-white road unevenness disturbance input. SAE Paper No. 902283, 1990.
131. Yamashita, M., K. Fujimoti, C. Uhlik, R. Kawatani, and H. Kumura, Hinf control of an automotive active suspension. *Proceedings of the 29th Conference on Decision and Control*, Honolulu, HI, 1990, pp. 2244–2250.
132. DeJager, A. G., Comparison of two methods for the design of active suspension systems. *Optimal Control Application*, 12, 173–138, 1991.
133. Balas, G. J., J. C. Doyle, K. Glover, A. Packard, and R. Smith, μ—Analysis and synthesis toolbox. *Matlab User's Guide*, MathWorks Inc., 1993.
134. Hrovat, D. and B. Bodenheimer., Robust automotive idle: Speed control design based on μ-synthesis. *Proceedings of the 1993 American Control Conference*, San Francisco, CA, June 1993, pp. 1778–1783.
135. Lu, J. and M. De Poyster, Multi-objective suspension control to achieve integrated ride and handling performance. *IEEE Transactions on Control Systems Technology*, 10(6), 807–821, 2002.
136. Gordon. T. J., C. Marsh, and M. G. Milsted, A comparison of adaptive LQG and nonlinear controllers for vehicle suspension systems. *Vehicle System Dynamics*, 20, 321–340, 1991.
137. Karlsson, N., M. Ricci, D. Hrovat, and M. Dahleh, A suboptimal nonlinear active suspension. *Proceedings of the 2000 American Control Conference*, Chicago, IL, 2000, pp. 4036–4040.
138. Alleyne, A., P. D. Neuhaus, and J. K. Hedrick, Application of nonlinear control theory to electronically controlled suspensions. *Vehicle System Dynamics*, 22, 309–320, 1993.
139. Gopalasamy, S., C. Osorio, J. K. Hedrick, R. Rajamani, for active suspensions—Controller design and experimental study. *ASME Journal of Dynamic Systems, Measurement and Control*, 61, 725–733, 1997.
140. Lin, J. S. and I. Kanellakopoulos, Nonlinear design of active suspensions. *IEEE Control Systems Magazine*, 17(3), 45–59, 1997.
141. Tran, M. and D. Hrovat, Application of gain-scheduling to design of active suspensions. *Proceedings of the 1993 IEEE Conference on Decision and Control*, San Antonio, TX, 1993, pp. 1030–1035.
142. Garcia, C. E., D. M. Prett, and M. Morari, Model predictive control: Theory and practice—A survey. *Automatica*, 25(3), 335–348, 1989.
143. Bemporad, A. and M. Morari, Control of systems integrating logic, dynamics and constraints. *Automatica*, 35(3), 407–427, 1999.
144. Borrelli, F., *Constrained Optimal Control of Linear and Hybrid Systems*. Vol. 290, Lecture Notes in Control and Information Sciences, Springer, Berlin, Germany, 2003.
145. Bemporad, A., F. Borrelli, and M. Morari, Model predictive control based on linear programming —The explicit solution. *IEEE Transactions on Automatic Control*, 47(12), 1974–1985, 2002.
146. Bemporad, A., M. Morari, V. Dua, and E. N. Pistikopoulos, The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1), 3–20, 2002.
147. Hrovat, D., MPC-based idle speed control design for internal combustion engines. *Proceedings of the 26th FISITA Congress*, Prague, Czech Republic, 1996.
148. Giorgetti, N., A. Bemporad, I. V. Kolmanivsky, and D. Hrovat, Explicit hybrid optimal control of direct injection stratified charge engines. *Proceedings of the International Symposium on Industrial Electronics (ISIE 2005)*, Mini Track on Automotive Control, Dubrovnik, Croatia, June 2005, pp. 247–252.

149. Petterson, S. and B. Lennartson, Stability analysis of hybrid systems—A gearbox application. In Johansson, R. and Rantzer, A. (eds.), *Nonlinear and Hybrid Systems in Automotive Control*, Springer, London, U.K., 2003, pp. 373–387.
150. Borrelli, F., A. Bemporad, M. Fodor, and D. Hrovat, An MPC/Hybrid system approach to traction control. Accepted for *IEEE Transaction on Control Systems Technology*. 14(3), 2006, pp. 541–552.
151. Isidori, A., *Nonlinear Control Systems*, 2nd edn., Springer-Verlag, Berlin, Germany, 1989, (1st Edn., 1985).
152. Astrom, K. J. and B. Wittenmark, *Adaptive Control*. Addison-Wesley, Reading, MA, 1989.
153. Slotine, J. J. E. and W. Li, *Applied Nonlinear Control*. Prentice-Hall, Englewood Cliffs, NJ, 1991.
154. Khalil, H.K., *Nonlinear Systems*. Macmillan, New York, 1992.
155. Krstic, M., I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*. John Wiley & Sons, New York, 1995.
156. Blankenship, G. L., R. Ghanadan, and V. Polyakov, Nonlinear adaptive control of active vehicle suspensions. *Proceedings of the 1993 American Control Conference*, San Francisco, CA, pp. 2837–2841.
157. Cheok, K. C., N. K. Loh, H. D. McGee, and T. F. Petit, Optimal model-following suspension with microcomputerized damping. *IEEE Transactions on Industrial Electronics*, IE-32, 364–371, 1985.
158. Sunwoo, M., K. C. Cheok, and N. J. Huang, Application of model reference adaptive control to active suspension systems. *Proceedings of the 1990 American Control Conference*, San Diego, CA, 1990, pp. 1340–1346.
159. Karlsson, N., D. Hrovat, and M. Dahleh, Nonlinear H_∞ control of active suspensions. *Proceedings of the 2001 American Control Conference*, Arlington, VA, 2001.
160. Jurgen, R., *Automotive Electronics Handbook*, Mc-Graw-Hill, New York, 1995.
161. Nagai, M. and Y. Sawada, Active suspension for flexible structure control of high speed ground vehicles. *Proceedings of the 10th IFAC World Congress*, Munich, Germany, 1987.
162. Gohring, E., R. Povel, E. C. von Glasner, and P. Schutzner, Practicable controlled suspension systems for commercial vehicles. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Yokohama, Japan, 1992, pp. 99–104.
163. Vohringer, K. D., O. Bode, E. C. von Glasner, H. Ch. Pflug, and R. Povel, Contribution to adaptive suspension systems for commercial vehicles. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Tsukuba, Japan, 1994, pp. 359–366.
164. Goodall, R. M., R. A. Williams, A. Lawton, and P. R. Harborough, Railway vehicle active suspensions in theory and practice. *Proceedings of the 7th IAVSD Symposium*, Cambridge, U.K., 1981, pp. 210–215.
165. Williams, R. A., Railway vehicle suspensions enter the electronic age. *IEE Conference*, London, U.K., 1981, p. 203.
166. Goodall, R. M. and W. Kortum, Active controls in ground transportation—A review of the state-of-the-art and future potential. *Vehicle System Dynamics*, 12, 225–257, 1983.
167. Gottzein, E., K. H. Brock, E. Schneider, and J. Pfefferl, Control aspects of a tracked magnetic levitation high speed test vehicle. *Automatica*, 13, 205–223, 1977.
168. Nagai, M. and S. Tanaka, Study of the dynamic stability of repulsive magnetic levitation systems (optimal control of active secondary suspension). *JSME International Journal*, Series 3, 35(1), 1992, pp. 102–108.
169. Horton. D. N. L. and D. A. Crolla, Theoretical analysis of a semi-active suspension fitted to an off-road vehicle. *Vehicle System Dynamics*, 15, 351–372, 1986.
170. Metz, D. and J. Maddock, Optimal ride height and pitch control for championship race car. *Automatica*, 22(5), 509–520, 1986.
171. Sorensen, A. J. and O. Egeland, Design of ride control system for surface effect ships using dissipative control. *Automatica*, 31(2), 183–199, 1995.
172. Yokoya, Y., R. Kizu, H. Kawaguchi, K. Ohashi, and H. Ohno, Integrated control system between active control suspension and four wheel steering for the 1989 CELICA. SAE Paper No. 901748, 1990.
173. Matsuo, Y., H. Harada, and M. Yamamoto, Development of experimental vehicle with integrated chassis control. *JSAE Review*, 11(3), 30–36, 1990.
174. Hrovat, D., Wheel slip control utilizing active suspension. US patent 5,033,573, 1991.
175. Hrovat, D., Traction control system with active suspension. US Patent 5,517,414, 1996.
176. Yonekawa, T., T. Ohnuma, V. Mon, T. Gotoh, and S. Buma, Effect of active control suspension system on vehicle dynamics. *JSAE Review*, 12(2), 40–45, 1991.
177. Fodor, M. and D. Hrovat, Vehicle suspension control. US Patent 5,852,787, 1998.
178. Karlsson, N., Constructive methods for nonlinear control of finite and infinite dimensional systems. PhD Thesis, University of California, Santa Barbara, CA, 2002.

32 Integrated Controls

Masato Abe

CONTENTS

32.1 Introduction 1227
32.2 Vehicle Chassis Control 1228
32.3 Feature of Individual Control 1229
32.4 Integrated Controls 1231
32.4.1 Integration of ABS/TCS and Steering Control 1231
32.4.2 Integration of Front and Rear Steer Controls 1232
32.4.3 Integration of Rear Wheel Steer and Front/Rear Torque Distribution of 4 WD 1233
32.5 Integration of Steer Control with Active Suspension/RDC 1233
32.6 Integration of Steer Control(s) and DYC 1235
32.7 Integrated Controls Aiming at Optimum Tire Force Distribution 1237
32.8 Conclusions 1242
Acknowledgment 1242
References 1242

32.1 INTRODUCTION

Vehicle motion controls in longitudinal, lateral, and vertical directions were to be controlled independently by traction/braking, steering, and suspension controls, respectively, in the past; however, in the recent 10–15 years, attentions are focused on integration of these controls to control vehicle horizontal motions, above all the vehicle lateral motion. We can call these recent years "integrated control era." In the early stage of the earlier-mentioned 10–15 years, the integration or system combination of vehicle motion control has been reviewed in [1] in which the integration possibilities of chassis and TCS then available are dealt with. There, integration is understood as functional or hardware integration or as functional and hardware integration.

Also in [2], a study is presented dealing with the synergism coming from the integration of such controlled systems which are designed for improving dynamic behavior of road vehicles. Several combinations of suspension, torque, and/or steering controls are presented and the idea is laid behind that integrated control is a combination of more than two controlled systems to offer synergistic effects on performance improvement in vehicle lateral dynamics.

Understanding the integrated control in this article is almost the same as in the earlier [1] and [2]. Control integration from the viewpoint of performance improvement of vehicle lateral dynamics by multiple hardware combinations that eventually result in the integral use of the tire forces will be dealt with here. So in other words, integrated control could be said as a reasonable use of the tire forces with several hardware and at the same time offering performance benefit of vehicle lateral motion.

It is interesting to see which areas are likely to be beneficial in a "g–g diagram" for respective system integrations in a conceptual diagram as shown in Figure 32.1 [3] which was also published in the early stage of the recent 10–15 years. The circular limit of longitudinal and lateral accelerations corresponds to tire friction circle for each tire and it is reasonable to regard the integrated vehicle control as the integral use of the tire forces.

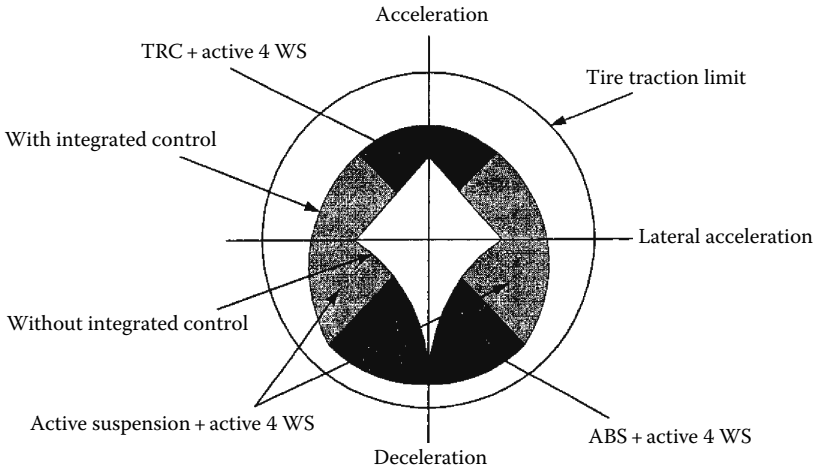


FIGURE 32.1 A concept of integrated control.

One of the important points in integrated control is a control integration methodology. Recently the author of [4] considers the scope, methodologies, and architectures for the design and development of interacting control systems in road vehicles laying an emphasis on the role of control system architecture, the resulting flow of control information, and the implications for control system design. It gives us broad review of current practice in integrated vehicle control to be referred once and so the earlier points will not be considered in any detail in this chapter.

32.2 VEHICLE CHASSIS CONTROL

Vehicle chassis control is generally meant to control vehicle lateral, vertical, and longitudinal motions in order to improve handling performance and safety, ride comfort, and traction/braking performance. These controls have basically been depending on steering controls, traction/braking controls, and suspension control, respectively. Our main interest here is the research and development of chassis controls for improving vehicle handling performance and active safety.

Since the 1980s, one of the most common chassis controls for the improvement of handling performance is rear-wheel steering (RWS). The RWS control depends on tire lateral force which is proportional to the steer angle, a control command, in a range where the lateral acceleration is small. In this range, the control law can rather easily be introduced by using two-degree-of-freedom linear vehicle model. However, in higher lateral acceleration, the lateral force is not necessarily proportional to the steer angle due to its saturation property to sideslip angle. The lateral force also becomes strongly dependent on the tire vertical load and the longitudinal force as well, so that the control law becomes rather sensitive to the vehicle motion and its environmental conditions [5].

In the recent 10–15 years, a so-called direct yaw moment control (DYC) is becoming one of the most promising means of chassis control. In DYC a vehicle motion is controlled by a yaw moment that is actively generated by the intentional distribution of the tire longitudinal forces between right and left (transversal split control of traction/braking). Some types of DYC are already installed in real cars in the market. One of the major advantages of this control method is that the tire longitudinal force has basically no feedback from the vehicle lateral motion as long as it is within the limit of the tire capacity due to the vertical load. One may think that the controlled longitudinal forces also have some effects of lateral motion, for example, effects of sideslip angle; however, it is not correct. A reason is as follows: A control command of the longitudinal force is normally a brake pressure or a traction torque which is directly an input torque to a spinning motion of a wheel.

The wheel spinning motion is independent from the vehicle lateral motion. So, the wheel longitudinal slip reaches some value at which the tire longitudinal force is balanced to the input torque, whatever a tire sideslip is, so far as the vertical load is large enough to produce the longitudinal force. This is a very important point in an intentional use of tire longitudinal forces for controlling vehicle lateral motion. Therefore, a precise yaw moment, required to control the vehicle lateral motion, can be developed and the vehicle motion with the control becomes more robust to fluctuation in the vehicle running condition and its environment [5,6]. Of course it should be noticed that the tire lateral force has the effect of the longitudinal force even though it is determined independently from the lateral force.

Though front-wheel active steer (FWS) has the same shortcoming aspects as RWS, recently FWS is also becoming one of the effective tools of the active chassis control together with developments of the steer-by-wire systems. However, the control command of both FWS and RWS is not the lateral force but the steering angle; therefore, the lateral forces generated by FWS and RWS have less accuracy and will be more susceptible to the vehicle motion because the tire lateral force has a nonlinear relationship to the sideslip angle and the vertical load. In contrast, as it is mentioned earlier, the yaw moment produced by DYC will be more precise; on the other hand, it cannot directly control the lateral motion since it does not use any lateral force directly for the control [6].

For these reasons, it is possible to realize more effective chassis control by integration or coordination of FWS, RWS, and DYC to compensate the drawback of each control. Roll moment distribution control (RDC) in an active suspension is also considered to be one of the tools to control the vehicle motion. A controlled four wheel drive (4WD) with differential control is one of the tools of chassis control as well. Of course, there are some other promising chassis controls than FWS, RWS, DYC, RDC, and controlled 4 WD such as antilock brake system (ABS) and traction control system (TCS). Since these controls are aiming at not controlling the vehicle motion directly but controlling the tire performance as a component control, they are not necessarily regarded as a control device for the strategic integrated vehicle motion controls but possible to accomplish their aims by stand-alone use; however, the drawbacks of ABS and TCS can be compensated by FWS, RWS, DYC, etc.

Since any general definition of "integrated control" is not established yet, the term is used here for combination or coordination/cooperation of two or three chassis controls for the vehicle handling and active safety, some of which are sometimes possible to be called multivariable control. Anyhow, it is obvious that any chassis control aiming at improving vehicle handling performance must rely on the tire lateral force and that the tire longitudinal force can also be used for the chassis control. These two forces each rely upon the vertical load and are interdependent. This relationship makes the vehicle dynamic characteristics very complex and nonlinear, and also it is no exaggeration to say that nonlinear characteristics of vehicle dynamics and its loss of stability can be due to nonlinear tire characteristics. Thus comprehensive study and deep understanding of tire and vehicle dynamics are required to develop more effective chassis control strategies as the integrated controls. It is desirable for the combination or coordination/cooperation of two or three chassis controls to compensate for the drawback of each stand-alone control due to nonlinear tire and vehicle characteristics.

32.3 FEATURE OF INDIVIDUAL CONTROL

Before going into detailed descriptions of the integrated controls, confirming the distinctive feature of typical chassis controls especially FWS, RWS, and DYC that are possible to be introduced into and compose promising integrated controls will be informative.

The vehicle response transfer functions calculated with a linear two-degree-of-freedom vehicle model of sideslip and yaw rate, to FWS, RWS, and DYC that will be produced by tire longitudinal forces, respectively, are shown in the Table 32.1. From this calculation, it is possible to understand the

TABLE 32.1
Linear Two Degree of Freedom Vehicle Responses to Active Chassis Control Inputs

Yaw Rate To

FWS	DYC	RWS
$\frac{r(s)}{\delta(s)} = G_\delta^r(0)\frac{1+T_r s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{r(s)}{M(s)} = G_M^r(0)\frac{1+T_{rs} s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{r(s)}{\delta_r(s)} = -G_{\delta r}^r(0)\frac{1+T_{rr} s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$
$G_\delta^r(0) = \frac{1}{1+AV^2}\frac{V}{l}$	$G_M^r(0) = \frac{V(K_f+K_r)}{2K_f K_r l^2(1+AV^2)}$	$G_{\delta r}^r(0) = \frac{1}{1+AV^2}\frac{V}{l}$
$T_r = \frac{ml_f V}{2lK_r}$	$T_{rs} = \frac{mV}{2(K_f+K_r)}$	$T_{rr} = \frac{ml_r V}{2lK_f}$

Sideslip Angle Acceleration To

FWS	DYC	RWS
$\frac{\beta(s)}{\delta(s)} = G_\delta^\beta(0)\frac{1+T_\beta s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{\beta(s)}{M(s)} = -G_M^\beta(0)\frac{1}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{\beta(s)}{\delta_r(s)} = G_{\delta r}^\beta(0)\frac{1+T_{\beta r} s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$
$G_\delta^\beta(0) = \frac{1-(m/2l)(l_f/l_r K_r)V^2}{1+AV^2}\frac{l_r}{l}$	$G_M^\beta(0) = \frac{mV^2+2(l_f K_f - l_r K_r)}{4K_f K_r l^2(1+AV^2)}$	$G_{\delta r}^\beta(0) = \frac{1+(m/2l)(l_r/l_f K_f)V^2}{1+AV^2}\frac{l_f}{l}$
$T_\beta = \frac{IV}{2ll_r K_r}\frac{1}{1-(m/2l)(l_f/l_r K_r)V^2}$		$T_{\beta r} = \frac{IV}{2ll_f K_f}\frac{1}{1+(m/2l)(l_r/l_f K_f)V^2}$
$\frac{\ddot{y}(s)}{\delta(s)} = G_\delta^{\ddot{y}}(0)\frac{1+T_{y1}s+T_{y1}s^2}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{\ddot{y}(s)}{M(s)} = G_M^{\ddot{y}}(0)\frac{1+T_y s}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$	$\frac{\ddot{y}(s)}{\delta_r(s)} = G_{\delta r}^{\ddot{y}}(0)\frac{1+T_{yr1}s+T_{yr2}s^2}{1+(2\zeta s/\omega_n)+(s^2/\omega_n^2)}$
$G_\delta^{\ddot{y}}(0) = VG_\delta^r(0)$	$G_M^{\ddot{y}}(0) = VG_M^r(0)$	$G_{\delta r}^{\ddot{y}}(0) = VG_{\delta r}^r(0)$
$T_{y1} = \frac{l_r}{V}$	$T_y = -\frac{(l_f K_f - l_r K_r)}{V(K_f+K_r)}$	$T_{yr1} = -\frac{l_f}{V}$
$T_{y2} = -\frac{I}{2lK_r}$		$T_{yr2} = -\frac{I}{2lK_f}$

K_f, front cornering stiffness; K_r, rear cornering stiffness; l_f, distance from center of gravity to front axle; l_r, distance from c.g. to rear axle; l, wheel base; m, vehicle mass; I, yaw moment of inertia; V, vehicle speed; A, stability factor.

basic nature of each control. It is understood that almost the same yaw rate response is available to any of the FWS, RWS, and DYC and there is no problem to control yaw rate by any of the FWS, RWS, and DYC. The sideslip response transfer function to FWS input has a positive zero; therefore, it is somehow difficult to have an exact model following control of sideslip angle by FWS. The DYC may not be suitable to control the lateral motion or acceleration, because there is a big delay in the lateral acceleration response to DYC input and also, from the practical point of view, the response gain to the moment input is very small compared with the responses to steer inputs, so impractically large input moment would be required for the control. Since the lateral acceleration transfer function to RWS has a positive zero, it is hard to have an exact lateral acceleration model following control by RWS.

There has been a detailed comparison between the effects of DYC and RWS [7–9]. Theoretically speaking, as long as the vehicle motion is considered in a form of two-degree-of-freedom vehicle plane model and the discussion is limited in model following controls of sideslip angle and yaw rate, the same control result is possible by DYC as is possible by RWS. For example, since both transfer functions of sideslip angle and yaw rate to yaw moment as well as to RWS have no positive zero, it is possible to introduce control laws of a sideslip angle or a yaw rate model following control by both DYC and RWS, respectively, just applying an inverse method. Even though it is

certain that RWS and DYC are possible to give us the same vehicle handling response, DYC is not necessarily suitable for the zero sideslip control. If the vehicle sideslip is made zero by the yaw moment produced by the tire longitudinal forces, the sideslip angle of rear wheels becomes smaller and the lateral force cannot sufficiently be generated at the rear wheels. In this case, the front wheels must take care of the lateral force to be balanced with the centrifugal force. Therefore, it will cause imbalance in the duty forces between the front and rear wheels that results in lowering the limitation of the vehicle motion.

The advantage of DYC is proved by the vehicle response to more severe steering input such as on a slippery road or severe lane change with braking. The vehicle motion with DYC still remains stable to such a severe maneuvering, while the vehicle motion with RWS starts spinning out. This is presumably due to the fact that the steering command from RWS control law cannot further generate sufficient lateral force of the rear wheels to control the vehicle motion to prevent the spin. Thus, the effect of DYC compared with that of RWS is more notable in the nonlinear range close to its limit. As long as the motion control depends only on the tire lateral force like RWS, there is a limit, because of the saturation characteristics, to the motion control with a large sideslip and lateral force. We can see that DYC has higher ability and is more effective even in such a near saturation region. Let's imagine a tire friction circle. At a near top or a bottom of the circle where the lateral force (vertical axis) is almost saturated, the margin of the longitudinal force (horizontal axis) is rather larger than that of the lateral force. At that region, using longitudinal force is more reasonable than using lateral force for the vehicle motion control.

The earlier characteristics are proved by using the so-called β-method [10,11] that is based on the relationships of restoring yaw moment to sideslip angle. This is a method to evaluate the vehicle dynamic characteristics by restoring yaw moment produced by sideslip motion of the vehicle. The restoring yaw moment decreases with braking and its excessive reduction makes the vehicle oversteer and causes vehicle spin. The moment increases with acceleration and excessive increase makes the vehicle too much understeer. The method compares how much an improvement can be made by the yaw moment produced by the tire longitudinal forces, that is, DYC and the rear tire lateral forces, that is, RWS, on the deterioration of yaw moment characteristics caused by the acceleration and deceleration. As a result, it is apparent that DYC is more effective and covers a wider range of the sideslip angle. Also the author of [12] has illustrated the effectiveness of DYC by using the moment method proposed by Milliken [13], that is, the relationship of yaw moment generated by tires against the lateral acceleration. He shows the result of the measurement of yaw moment generated against the vehicle lateral acceleration for vehicles equipped with DYC and with RWS using a flat belt-type test machine. The results show that the vehicle with DYC maintains its control up to higher lateral acceleration range.

32.4 INTEGRATED CONTROLS

32.4.1 Integration of ABS/TCS and Steering Control

As is described in Section 32.2, ABS and TCS are not necessarily regarded as a control device for the strategic integrated vehicle motion controls; however, the ill effects of ABS and TCS on vehicle yaw motion can be compensated by FWS, RWS, etc., and [2] deal with the synergism of the combination of ABS with yaw-controlled four wheel steering (4WS). Though ABS can provide to avoid the wheel lock and maintain the tire ability to produce the lateral force during braking, all the wheels of the vehicle cannot be controlled individually because the difference of tire longitudinal forces between right and left especially during braking on μ-split road surface impairs the directional stability. Therefore, the wheels are controlled dependently which results in a longer stopping distance. [2] shows that the combined use of the yaw-controlled 4 WS with the individually controlled ABS can give us shorter stopping distance as well as higher directional stability.

32.4.2 Integration of Front and Rear Steer Controls

As an extension of RWS, there is a concept of the integration of front and rear steer controls. By this combined use of the controls, as there are two control variables, it is obvious that the exact model following control of, for example, sideslip/lateral acceleration and yaw rate simultaneously is possible; however, there is a common shortcoming in FWS and RWS and so much effect of the integration may not be expected.

There is a trial to apply a model matching control by linear quadratic (LQ) method for integrating FWS and RWS controls [14]. In this concept of the controls, the vehicle model used to introduce the control is linear and there is no intentional consideration to solve the problems due to tire nonlinear characteristics, however, the effects of the integrated control that is eventually extended to nonlinear region are found. The control concept can be stated as a simultaneous use of two controls (multivariable variable control) for the model matching control of sideslip and yaw motions.

In order to intentionally cope with the problem due to tire nonlinear characteristics, there is an application of two-degree-of-freedom/H-infinity control to the synthesis of integrated controller of FWS and RWS though the vehicle model used is a two-degree-of-freedom linear one [15]. The controller consists of a feed-forward part and feedback one. Basically feed-forward controller is designed for the vehicle to follow the model responses of lateral acceleration and yaw rate. The feedback controller is designed to improve the robustness of the system against changes in vehicle dynamics presumably due to nonlinear tire characteristics, applying the H-infinity control system design. The frequency responses of an experimental vehicle with the integrated control on a wet road are shown in Figure 32.2. Even maneuvering on a wet road surface, the responses of the vehicle agree with those of the reference model over wide frequency range. Of course the vehicle response characteristics are significantly improved by the integrated control compared with the vehicle without the integrated control. This is recognized in the shape of the frequency response in Figure 32.2 that is very close to the reference model response and fairly far from the well-known lateral acceleration and yaw rate frequency response curves of the conventional vehicle even under wet conditions (Figure 32.2).

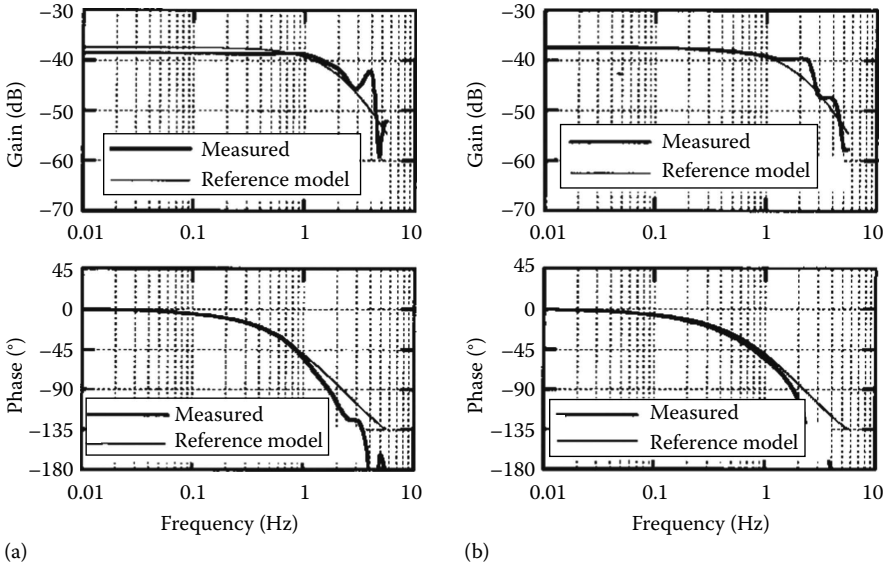


FIGURE 32.2 Integration effects of front and rear active steers by two-degree-of-freedom/H-infinity control. (a) Lateral acceleration and (b) yaw rate. (From Horiuchi, S. et al., *Vehicle Syst. Dynamics*, 25, 275, 1996.)

32.4.3 Integration of Rear Wheel Steer and Front/Rear Torque Distribution of 4 WD

Though the longitudinal force of the tire gives the indirect but significant effects on the tire lateral force, a torque distribution control by differential control of 4 WD cannot directly produce the tire lateral forces. Even an arbitrary yaw moment cannot be produced just by the differential controls. So the control cannot be used for any response model following control and there are few examples of integral use of the control with the other controls.

Even so, there is an application of nonlinear robust control to introducing the integrated control of RWS and torque distribution of 4 WD [16,17]. The control is based on a "two-degree-of-freedom" control system. The feed-forward part is composed only of RWS by solving the linear inverse problem to follow the yaw rate model response. The objective of the feedback part is to stabilize the vehicle in nonlinear regions and is therefore based primarily upon 4 WD control because the effect of the longitudinal force controlled by 4 WD torque distribution on tire characteristics itself is quite significant in nonlinear region. The RWS is also used in the feedback compensator in the linear region. A simple nonlinear tire model, in which the saturation property of the lateral force to sideslip angle and the effects of the longitudinal force on reducing the lateral force of the tire are considered, is used in introducing the feedback control. A strict linearization method is adopted otherwise; as the provided vehicle model is nonlinear, it is impossible to apply a linear control theory to determine the rear wheel steer angle and the front/rear torque distribution ratio. The feed-forward control part of the rear wheel steer in both [16] and [17] is designed to follow a given yaw rate on one hand; H-infinity control in [16] and μ-synthesis in [17] are used to design the feedback compensator, respectively. The effects of the integrated control are prominent as is expected especially in tire nonlinear regions.

32.5 INTEGRATION OF STEER CONTROL WITH ACTIVE SUSPENSION/RDC

A primary objective of active suspension is to improve ride comfort and it does not directly produce lateral forces so that it cannot be used alone to intentionally control the vehicle lateral motion or cannot necessarily be regarded as a strategic device for an integration control with other controls. Nonetheless, there is an example to deal with the synergism between active suspension and 4 WS in a straight running on irregular profile surface roads. Though it is just the combined use, effects are significant. Details are presented in [2] and will not be presented in any detail here.

A balance of the front/rear roll stiffness has a significant effect on vehicle understeer/oversteer characteristics depending on the transversal load transfer during circular turning which reduces the sum of the lateral forces of right and left wheels due to nonlinear relation of tire lateral force to vertical load. Hence, the RDC in active suspension or roll stiffness distribution control, RDC, is regarded as one of the tools for the chassis control; however, the RDC cannot directly produce tire lateral force so that the control cannot be used for any type of model following control as is for a conventional type of controlled 4 WD, and so there are few examples of integral use of the control with the other controls.

There is an example of just an intentionally combined use of RDC with RWS [18]. It is obvious that any type of RWS is inevitable to have the limit due to saturation property of the rear tire that causes the deterioration of the vehicle stability because the rear wheels of the vehicle with RWS must have higher tire work load. Depending upon an intuitive understanding that reducing the roll moment distribution at the rear axle is expected to mitigate this condition, the RDC is used aiming at compensating the drawback of the vehicle with RWS. The roll moment distribution in this example is controlled referring a spin or a drift level measured by the sideslip velocities at front and rear ends of the vehicle. If the driver's will is to increase the yaw rate and the vehicle motion has a tendency toward drift—front sideslip velocity is larger than rear one—then the roll moment is distributed to the rear. If the driver's will is to reduce the yaw rate and the tendency of the vehicle motion is toward spinning—rear sideslip velocity is larger

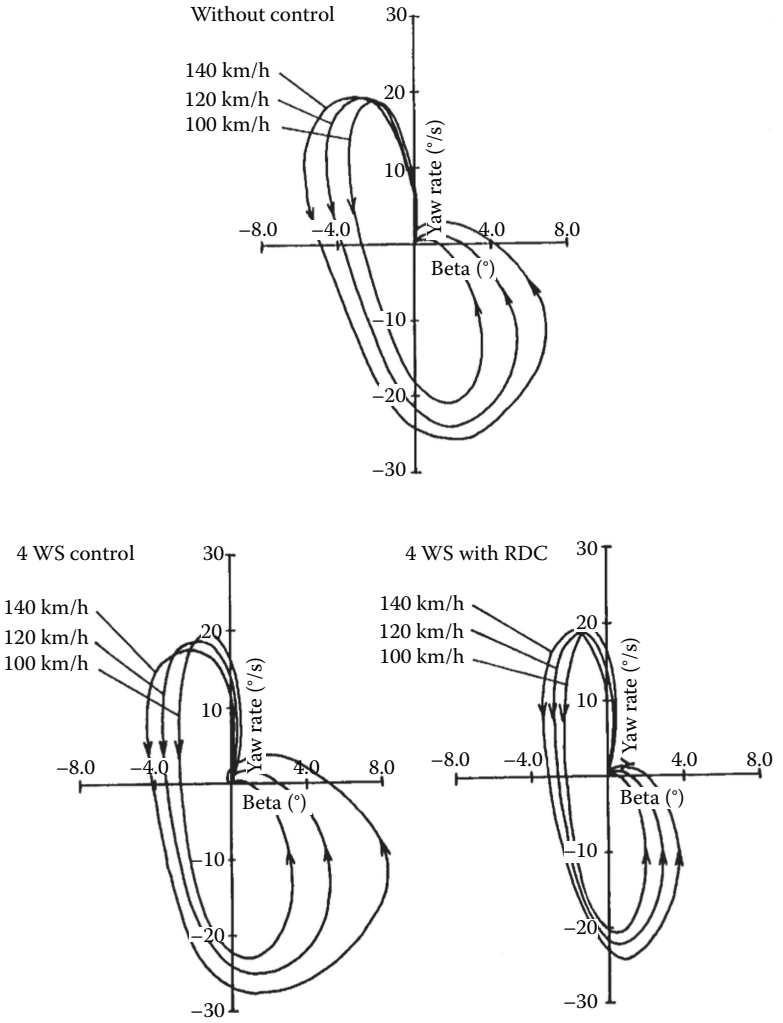


FIGURE 32.3 Effects of the integration of RDC and 4 WS. (From Abe, M., *JSAE Rev.*, 12(3), 42, 1991.)

than front one—then the roll moment is distributed to the front. The control law of the RWS considered here is that the rear wheel is steered proportionally to the front steer angle so that the steady-state sideslip angle equals zero. The results are shown in Figure 32.3 and summarized by the sideslip and yaw rate response state plane to single sine-wave steer input. It is found that the RWS alone does not necessarily give the vehicle motion with smaller sideslip or reduce spin tendencies in the vehicle motion with high lateral acceleration. It is obvious, however, that the combination control with the RDC can eliminate large sideslip angle and reduce the spinning tendency at the latter part of the single sine-wave steer input, though the stand-alone use of the 4 WS brings us the vehicle motion more liable to spin at high speed rather than even the conventional vehicle without the 4 WS.

Reference [19] presents more strategic and theoretically more sophisticated examples of the integrated controls of the steer controls and RDC. Three control input variables—front and rear steer angles and roll moment distribution ratio—are adopted and the nonlinear vehicle model is provided especially considering tire nonlinear characteristics. In order to apply a linear control theory, the vehicle model is exactly linearized by nonlinear transformation from three input variables to two pseudo-inputs, in which the linearized vehicle includes perturbations in the front and

rear cornering stiffness. The control law of two pseudo-inputs is designed using "two degree of freedom" control system and the feed-forward part is designed by model matching method; on the other hand, feedback part is designed by applying μ-synthesis. The roll moment distribution ratio is determined to minimize the square sum of the front and rear tire sideslip angles and the difference of the distribution ratio from 50/50, in which both sideslip angles are functions of predetermined two pseudo-input variables and the roll moment distribution ratio. The effects of the integrated control are prominent especially in tire nonlinear region under low-friction road surface.

Reference [20] is an additional example of combined use of RDC with an active drivetrain for variable torque distribution at a center differential, a kind of controlled 4 WD. The RDC controller is a yaw rate model reference PI controller and a variable torque distribution control is based on a sideslip angle feedback PI controller. In the combined control, the center differential just distributes the torque to front and rear with the same ratio as the roll moment distributed by RDC; however, the benefits of the combined control as well as the insight behind the integration strategy are not necessarily clear.

32.6 INTEGRATION OF STEER CONTROL(S) AND DYC

As is discussed in Section 32.2, not only steer controls (FWS and RWS) but also DYC is one of the promising components for the integrated controls, because it is a more effective vehicle control in nonlinear regions. Hence, there are many approaches to the integration of steer control(s) and DYC.

There are trials to apply a model matching control by LQ method for integrating steer control(s) and DYC [21–23]. They are simultaneous use of two or three control variables (multivariable control) for the model matching control. As is the same in the case of the introduction of the integration of steer controls—FWS and RWS in Section 32.4.2—the vehicle models used to introduce the controls are linear and no consideration to intentionally deal with the problems due to tire nonlinear characteristics is found in this strategy of the combined control. However, again the effect of the integrated controls that is extended eventually under nonlinear region is found.

In order to intentionally overcome the problem due to tire nonlinear characteristics, there are applications of the nonlinear predictive controller design for the reference model following control to introducing the integration of steer control(s) and a full DYC [24,25]. A nonlinear nominal vehicle model is assumed from the beginning in this controller design, and the theory for continuous time nonlinear predictive control is used for designing the integrated control. The effects of the integration are investigated by the simulations. An example of the simulation results is shown in Figures 32.4a and b in which it is clear that the vehicle response with the integrated control is closest to the reference model response in yaw rate as well as in sideslip angle compared with the responses with other controls.

As is described in Section 32.2, it is obvious that the RWS is effective rather in a linear range of vehicle motion; on the other hand, the DYC is more effective in a nonlinear range so that there is a simple and intuitive approach to the coordination strategy of FWS and DYC [26,27]. This approach in coordinating the two controllers is that FWS is used to improve the vehicle handling properties at low- to mid-lateral accelerations and the DYC is only used when the stability limit of the vehicle is approached. Both controllers are designed based on stand-alone controller design and a rule-based integration scheme is developed to coordinate the two stand-alone control systems. The rules are the product of investigations carried out empirically in simulation and are to distribute the control demands to the two individual controls. Though the integration strategy could be reasonable, the substantial effects are not necessarily clear so far.

Another integration approach to deal with the problems due to tire/vehicle nonlinear characteristics and compensate for drawbacks of the stand-alone using each active control is to

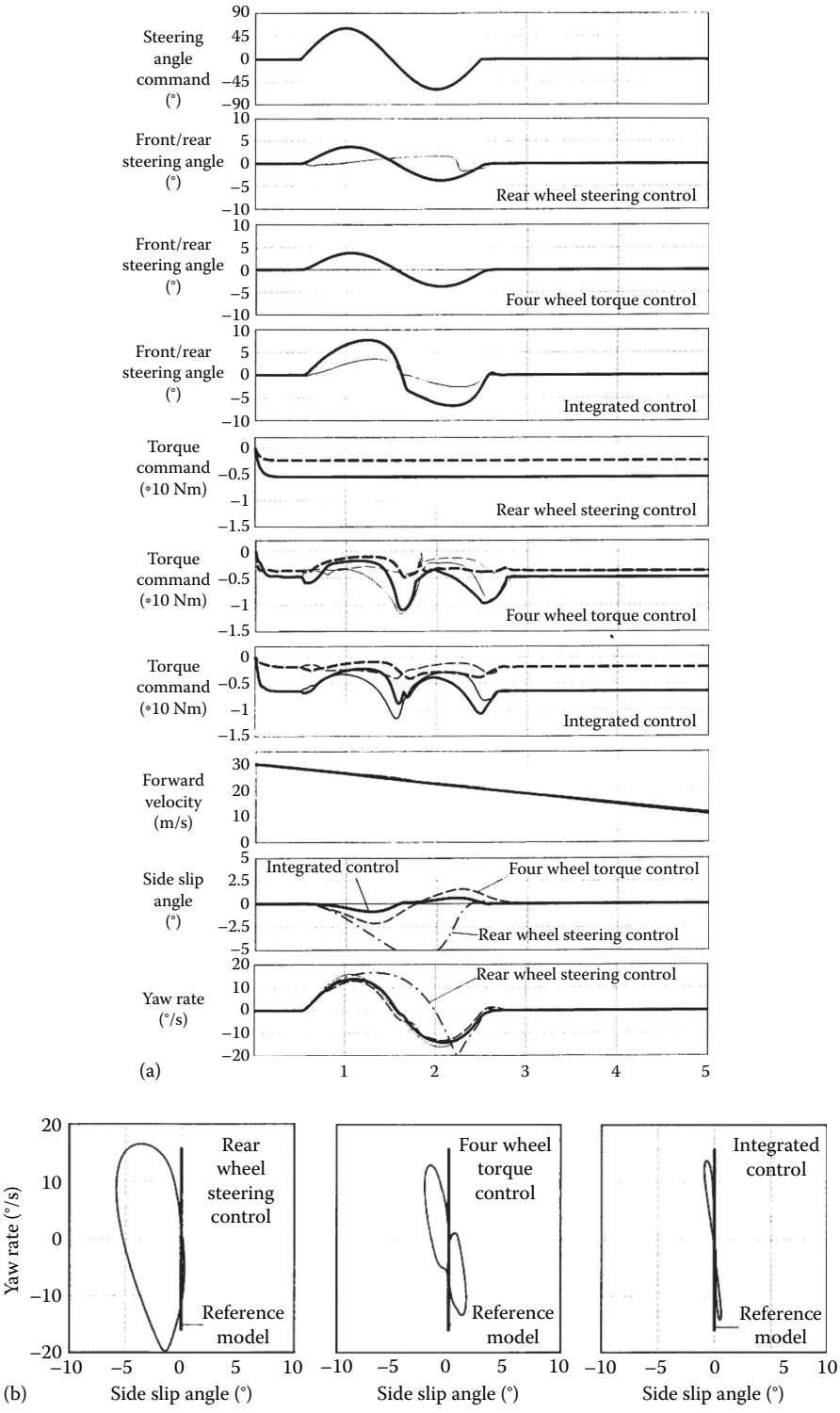


FIGURE 32.4 (a) Effects of the integrated controls on vehicle responses to sine-wave steer input with braking. (b) Effects of the integrated controls on vehicle response on state plane. (From Horiuchi, S. et al., *Vehicle Syst. Dynamics*, 33, 680, 1999.)

apply a sliding control type of nonlinear model following control to integrating steer control(s) and DYC [28–30]. A simple nonlinear tire model in which the saturation characteristics of lateral force to sideslip angle as well as the effects of longitudinal force on reducing lateral force with friction circle concept and the nonlinear dependence of lateral force on vertical load are considered appropriately is used in the controller design. As the intended integrated control has more than two control variables, the exact model following control of, for example, sideslip and yaw rate simultaneously is possible. In this approach, the sideslip and yaw rate responses of the two-degree-of-freedom vehicle model with linear tires are adopted as reference model responses.

Then it is expected that basically no active control is required so long as the vehicle motion remains in a linear range so that the primary objective of the integrated control eventually turns out to keep the vehicle response in a linear region as if it were monitoring how much the vehicle motion is nonlinear at each moment by using the simple nonlinear tire model. The model following error in sideslip angle is taken as a sliding surface and the first-order lag of the error equal to zero is taken as a sliding condition. From this condition it is possible to calculate a total lateral force required to follow the model sideslip response using the fact that the lateral force required is a product of vehicle mass and the required lateral acceleration.

The model following error in yaw rate is also taken as another sliding target and the first-order lag of the error equal to zero is again taken as a sliding condition. From this condition the total yaw moment required to follow the model yaw rate is led because the first-order derivative of the required yaw rate is the yaw moment required.

By using the simple tire model mentioned earlier, the active front and/or rear steer angle(s) to satisfy the calculated total lateral force and the yaw moment to be produced by longitudinal tire forces, DYC, is calculated so that the total yaw moment to be produced by tires is equal to the required one. In the case of the integrated control by FWS, RWS, and DYC, there are three control variables available and at least two control variables are enough to follow the model responses of sideslip and yaw rate. Therefore, the additional condition is added to share the total tire lateral force required between front and rear wheels proportionally to the vertical load of the tires in the controller design. At the same time, sideslip angle of the vehicle to be estimated by using the simple tire model is needed for the controller design because a direct measurement of it is difficult.

The effects of the integrated controls are proved by simulations with the full vehicle model. Responses to single sine steer input with braking are shown on the state plane in Figure 32.5 in which both integration controls of DYC + FWS + RWS and DYC + FWS show almost the same good results. On the other hand, if the model response of the sideslip angle is changed to the response that is the same response as the normal linear vehicle model at transient period converges to zero at steady state, then the results are changed as shown in Figure 32.6. The results show that the RWS is essential in the integrated control if the reduction of sideslip response, sometimes equal to zero, is required. You can refer back again to Table 32.1 on how the sideslip responses are to FWS, RWS, and DYC, respectively.

Reference [31] is another type of integration of RWS and DYC (or VDC) with roll control to be referred in which H-infinity integration as well as an intuitive integrated chassis design are dealt with.

32.7 INTEGRATED CONTROLS AIMING AT OPTIMUM TIRE FORCE DISTRIBUTION

The integrated controls will have multiplevariables for the control so that, as is described earlier, the target vehicle motion or performance is rather easier to be obtained by, for example, any type of possible model following control. Therefore, the problem is rather how much workload the tires have during working to keep the vehicle behave as given model response. Then the motivation arises to use each tire reasonably within a limit in the integrated control, which is expected to result

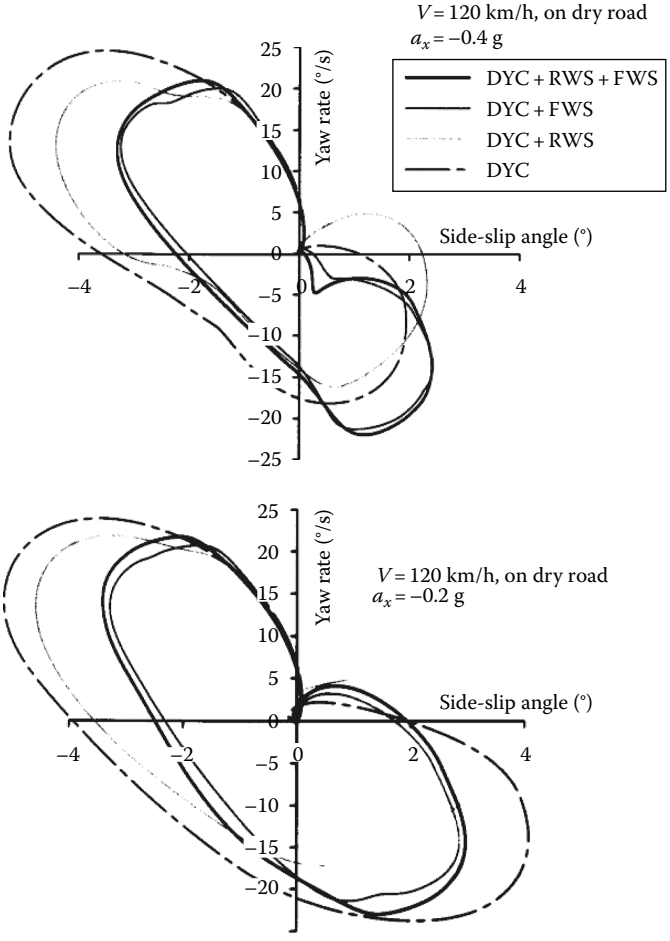


FIGURE 32.5 Effects of the integrated control on vehicle responses on state plane. (From Mokhiamar, O. and Abe, M., *Vehicle Syst. Dynamics*, 37, 246, 2002; Mokhiamar, O. and Abe, M., *Proc. IMechE Part D J. Automobile Eng.*, 216, 115, 2002.)

in solving the problems due to tire nonlinear characteristics and compensating the drawback of the stand-alone use of the chassis controls. It could be said that one of the generic strategies of the integrated controls is to distribute lateral and longitudinal forces to each tire to optimize the tire performance and obtain the required vehicle response. Of course the earlier idea rests on the premise of full drive-by-wire vehicle in which all the wheel steer angles and wheel torques are controlled independently to distribute lateral and longitudinal forces to each tire.

It is possible to mention the two important contributions [32,33] to the integrated controls in terms of the earlier text. An estimation of tire grip margin that is equivalent to friction coefficient between tire and road surface sensed the self-aligning torque of each tire is the basis of the control strategy in [33]. Based on the analysis of a physical tire model, it is assumed that a ratio of the measured self-aligning torque to a virtual self-aligning torque proportional to sideslip angle is equivalent to the margin of the tire lateral force. The following definition of the tire workload rate is adopted:

$$\gamma_i = \frac{\sqrt{X_i^2 + Y_i^2}}{\mu_i Z_i} \tag{32.1}$$

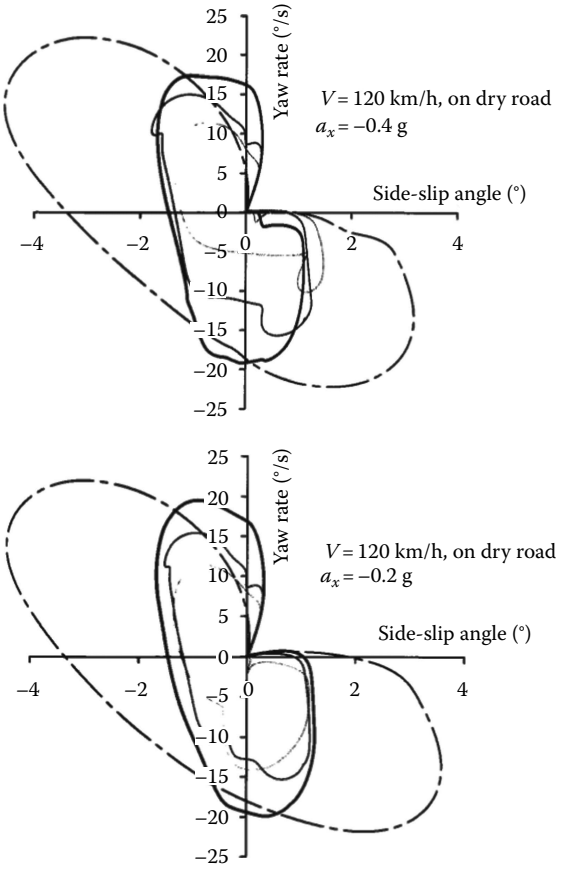


FIGURE 32.6 Effects of the integrated control of zero sideslip model response on vehicle motion. (From Mokhiamar, O. and Abe, M., *Vehicle Syst. Dynamics*, 37, 246, 2002; Mokhiamar, O. and Abe, M., *Proc. IMechE Part D J. Automobile Eng.*, 216, 115, 2002.)

where

X_i, Y_i, and Z_i (i = 1–4) are longitudinal and lateral forces and vertical load of each tire, respectively

μ_i is the friction coefficient between tire and road surface

A kind of nonlinear optimization technique, SQP (Sequential Quadratic Programming), provides the optimum distribution of the tire forces by keeping each γ_i to be equal and minimizing it under the constraints that the total lateral and longitudinal forces and the total yaw moment produced by tires satisfy the required target force and moment of the vehicle motion such as mentioned in Section 32.4.5. Once the lateral and longitudinal forces of each tire are determined, the steering and the traction or braking command are applied to each tire actuator to produce them. The effects of the integrated control are proved with the computer simulation. Additionally experimental investigations are carried out to validate the estimation of tire grip margin on icy road and also to prove the effects of the control applying only to the best balancing brake control of four wheels.

Other than the earlier, one of the possible approaches to the optimization of the tire performance is to minimize the square sum of tire workload expressed by

$$J = \sum_{i=1}^{4} C_i \frac{X_i^2 + Y_i^2}{Z_i^2} \tag{32.2}$$

where, C_i is a weighting coefficient and normally set equal to 1.0 [34,35]. This criterion of the tire performance optimization is basically the same as is adopted as Equation 32.1 in [33].

In order to guarantee the vehicle response to follow the model response described, for example, in Section 32.4.5 or whatever it is, a total lateral force of the tire must be equal to the target value to follow the sideslip model response and a total yaw moment produced by the tire forces must be equal to the target value to follow the yaw rate response. In addition to this, a total longitudinal force of the tire must be satisfied to reflect the driver's will of longitudinal speed control that is translated from brake or acceleration pedal pressure/stroke to longitudinal acceleration requirement. They are regarded as three constraints among the variables for the tire performance optimization and J is an objective function of the eight variables, X_i and Y_i (i = 1–4); thus, finally we can have the optimization problem with five variables without constraint. Applying the minimizing conditions, which are partial derivatives of J to the five variables equal to zero, to the objective function, it is easy to get a simple form of linear algebraic equation with five variables. Because the objective function is a second-order equation of the variables, the partial derivative of the function becomes always linear functions.

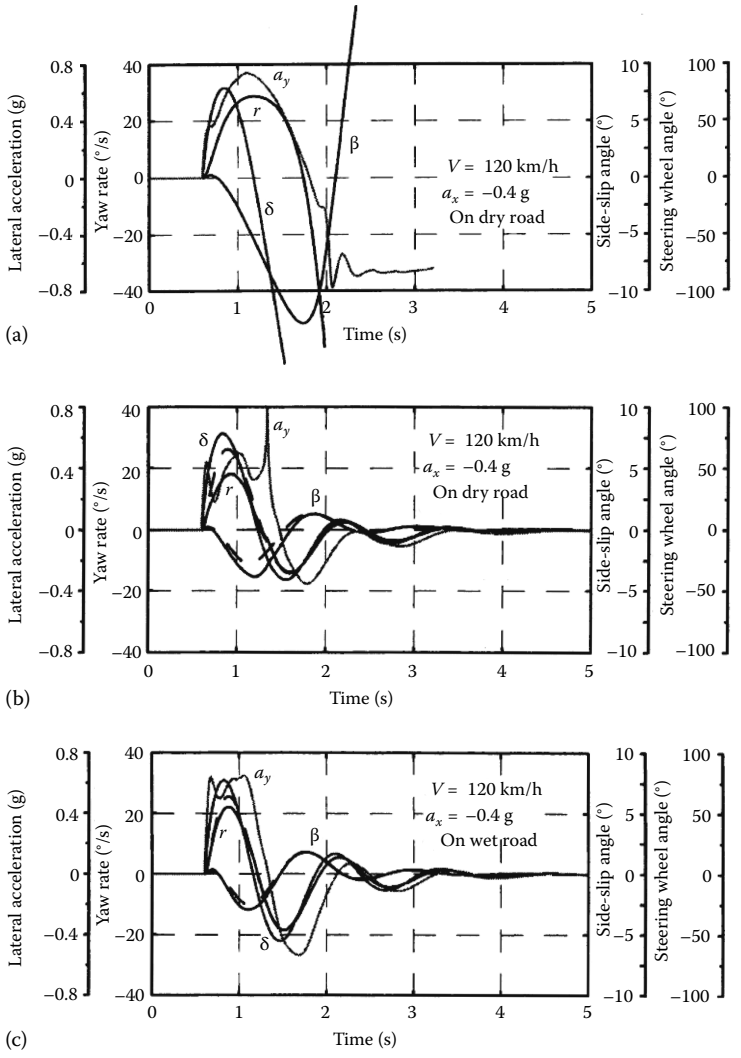


FIGURE 32.7 Effects of the integrated control on evasive lane change performance on dry road (a) Without control, (b) DYC + FWS + RWS, (c) Optimum tire force distribution. (From Mokhiamar, O. and Abe, M., *Trans. ASME, J. Dynamic Syst. Meas. Cont.*, 126, 753, 2004.)

Solving this algebraic equation at each time step brings us the optimum tire lateral and longitudinal forces to be distributed to each tire. This approach to the integrated control rests on the premise of full drive-by-wire system in which all the wheel steer angles and wheel torques are controlled independently. Thus once X_i and Y_i are determined, a torque of each wheel is given by $T_i = R_i X_i$ (R_i is a wheel radius) and a steering angle of each wheel δ_i is determined by using the simple nonlinear tire model mentioned in Section 32.4.5.

A computer simulation of the evasive lane change has validated the significant effects of the integrated control. The results are compared with the case of the integrated control of DYC + FWS + RWS mentioned in 32.4.5 and shown in Figures 32.7 and 32.8. The effect of the integrated control proposed is quite prominent especially in severe maneuvering conditions. However, the effects of the control on μ-split road surface are behind the effects brought by the integrated control introduced in [33] because the friction coefficient (tire grip margin) estimation is not taken into account [36]. Also it should be pointed out that it is possible to regard Z_i ($i = 1$–4) in the objective function J as variables to be controlled by the RDC with some constraints, though the problem becomes a strongly nonlinear optimization problem.

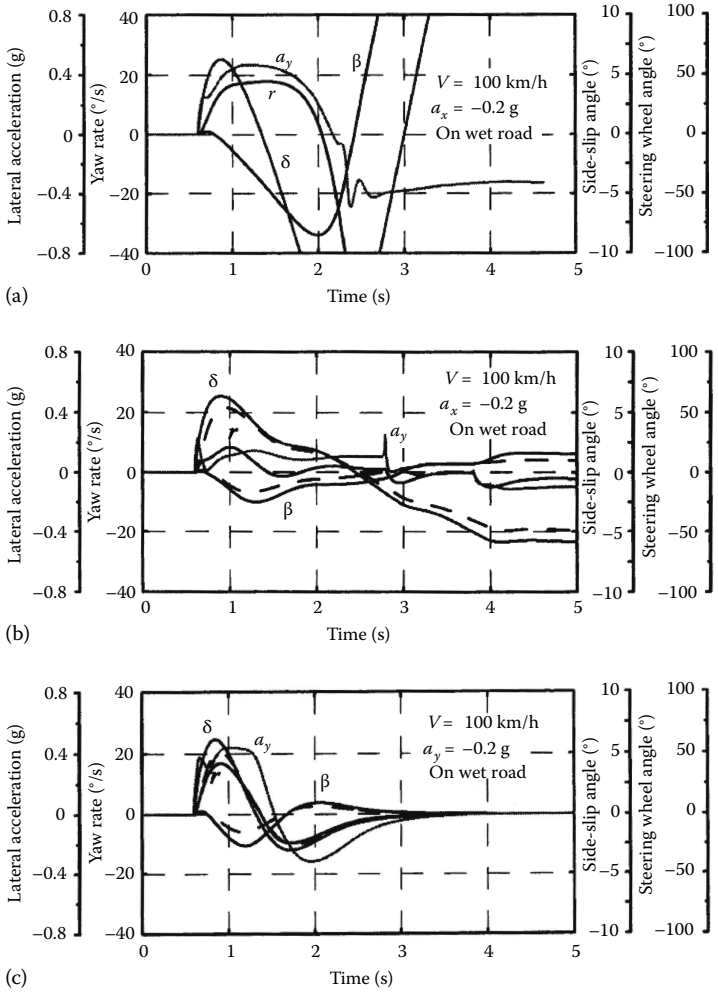


FIGURE 32.8 Effects of the integrated control on evasive lane change on wet road (a) Without control, (b) DYC + FWS + RWS, (c) Optimum tire force distribution. (From Mokhiamar, O. and Abe, M., *Trans. ASME, J. Dynamic Syst, Meas. Cont.*, 126, 753, 2004.)

32.8 CONCLUSIONS

A recent trend of the integrated controls is surveyed. An emphasis is laid on that the combination or coordination/cooperation of two or three chassis controls is expected to deal with the problems due to nonlinear tire and vehicle characteristics and to compensate the drawback of the stand-alone use of each control. Putting this viewpoint at the center, some details of typical integrated controls have been presented. It could be said that a future target of full drive-by-wire system is an ultimate integrated vehicle motion control system and the optimum use of the tires for getting required vehicle motion is the top priority for the integrated control. It is a control using tire forces in three directions arbitrarily for vehicle longitudinal, lateral, and vertical motion control depending on driver's three maneuvering commands *running, curving, and stopping.* A further research and development focusing on this direction especially from the practical point of view is recommended.

ACKNOWLEDGMENT

The author is indebted to all the authors in the following references.

REFERENCES

1. H. Wallentowitz, Integration of chassis and traction control systems: What is possible—What makes sense—What is under development, In *Proceedings of the International Symposium on Advanced Vehicle Control* (*AVEC92*), pp. 1–7, Yokohama, Japan, 1992.
2. G. Mastinu, E. Babbel, P. Lugner, D. Margolis, D. Mittermayr, and B. Richiter, Integrated controls of lateral vehicle dynamics, *Vehicle System Dynamics*, 23, 58–377, 1994.
3. S. Sato, H. Inoue, M. Tabata, and S. Inagaki, Integrated chassis control system for improved vehicle dynamics, In *Proceedings of the International Symposium on Advanced Vehicle Control* (*AVEC92*), pp. 413–418, Yokohama, Japan, 1992.
4. T. Gordon, M. Howell, and F. Brandao, Integrated control methodologies for road vehicles, *Vehicle System Dynamics*, 40, 157–187, 2003.
5. Y. Furukawa and M. Abe, Advanced chassis control systems for vehicle handling and active safety, *Vehicle System Dynamics*, 28, 59–86, 1997.
6. M. Abe, Vehicle dynamics and control for improving handling and active safety: From four-wheel steering to direct yaw moment control, *Proceeding of Institution of Mechanical Engineering, Part K: Journal of Multi-body Dynamics*, 213, 87–101, 1999.
7. M. Abe et al., Comparison of 4WS and direct yaw moment control (DYC) for improving of vehicle handling performance, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 159–164, Tsukuba, Japan, 1994.
8. M. Abe et al., A direct yaw moment control for improving limit performance of vehicle handling-comparison and cooperation with 4WS, *Vehicle System Dynamics*, 25, 3–23, 1996.
9. M. Abe et al., An experimental validation of side-slip control to compensate vehicle lateral dynamics for loss of stability due to nonlinear tire characteristics, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 179–186, Ann Arbor, MI, 2000.
10. Y. Shibahata et al., The improvement of vehicle maneuverability by direct yaw moment control, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 452–457, Yokohama, Japan, 1992.
11. K. Shimada et al., Comparison of three active chassis control methods for stabilizing yaw moments, SAE 940870, pp. 1178–1187, 1994.
12. S. Motoyama et al., Study on quantitative evaluation method for yawing control potential, *Journal of Society of Automotive Engineering of Japan*, 50(11), 52–57, 1996.
13. W. F. Milliken Jr. et al., The static directional stability and control, SAE Paper 760713, 1976.
14. M. Nagai, Active four-wheel-steering system by model following control, *Vehicle System Dynamics*, 18, 428–439, 1989.
15. S. Horiuchi et al., Two degree of freedom/H-infinity controller synthesis for active four wheel steering vehicles, *Vehicle System Dynamics*, 25, 275–292, 1996.
16. Y. Hirano et al., Integrated control system of 4WS and 4WD by H control, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 419–423, Yokohama, Japan, 1992.

17. Y. Hirano and O. Eiichi, Non-linear robust control for an integrated system of 4 WS and 4 WD, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 147–152, Tsukuba, Japan, 1994.
18. M. Abe, A study on effects of roll moment distribution control in active suspension on improvement of limit performance of vehicle handling, *Journal of Society of Automotive Engineering Review*, 12(3), 42–47, 1991.
19. E. Ono, Y. Hayashi, S. Doi, and K. Takanami, Coordination of vehicle steering and suspension systems by integrated control strategy, In *Proceedings of the International Symposium on Advanced Vehicle Control* (*AVEC92*), pp. 384–389, Yokohama, Japan, 1992.
20. N. Cooper et al., Study of the integration of roll control and torque distribution, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 513–518, Arnhem, the Netherlands, 2004.
21. M. Nagai et al., Integrated control law of active rear wheel steering and direct yaw moment control, *Vehicle System Dynamics*, 27, 357–370, 1997.
22. M. Nagai et al., Study on integrated control of active front steer angle and direct yaw moment, *Journal of Society of Automotive Engineering Review*, 23(3), 309–315, 2002.
23. M. Shino et al., Vehicle handling and stability control by integrated control of direct yaw moment and active steering, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 25–31, Hiroshima, Japan, 2002.
24. S. Horiuchi et al., Integrated control of four wheel steering and wheel torques using nonlinear predictive controller, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 111–116, Nagoya, Japan, 1998.
25. S. Horiuchi et al., Effects of integrated control of active four wheel steering and individual wheel torque on vehicle handling and stability—A comparison of alternative control strategies, *Vehicle System Dynamics*, 33, 680–691, 1999.
26. M. Selby et al., A coordination approach for DYC and active front steering, SAE2001–01–1275.
27. J. He et al., Integrated chassis control through coordination of active front steering and intelligent torque distribution, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 333–339, Arnhem, the Netherlands, 2004.
28. Y. Furukawa and M. Abe, On-board-tire-model reference control for cooperation of 4WS and direct yaw moment control for improving active safety of vehicle handling, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 507–527, Aachen, Germany, 1996.
29. O. Mokhiamar and M. Abe, Combined lateral force and yaw moment control to maximize stability as well as vehicle responsiveness during evasive maneuvering for active vehicle handling safety, *Vehicle System Dynamics*, 37, 246–256, 2002.
30. O. Mokhiamar and M. Abe, Active wheel steering and yaw moment control combination to maximize stability as well as vehicle responsiveness during quick lane change for active vehicle handling safety, *Proceedings of the Institution of Mechanical Engineering, Part D: Journal of Automobile Engineering*, 216, 115–123, 2002.
31. K. Kitajima et al., H infinity control for integrated side-slip, roll and yaw controls for ground vehicles, In *Proceeding of the International Symposium on Advanced Vehicle Control*, Ann Arbor, MI, 187–194, 2000.
32. Y. Hattori et al., Force and moment control with nonlinear optimum distribution for vehicle dynamics, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 595–600, Hiroshima, Japan, 2002.
33. E. Ono et al., Vehicle dynamics control based on tire grip margin, In *Proceeding of the International Symposium on Advanced Vehicle Control*, pp. 531–536, Arnhem, the Netherlands, 2004.
34. O. Mokhiamar and M. Abe, Simultaneous optimal distribution of lateral and longitudinal tire forces for the model following control, *Transactions of the American Society of Mechanical Engineering, Journal of Dynamic Systems, Measurement and Control*, 126, 753–763, December 2004.
35. O. Mokhiamar and M. Abe, Experimental verification using a driving simulator of the effect of simultaneous optimal distribution of tire forces for active vehicle handling control, *Proceedings of the Institution of Mechanical Engineering, Part D: Journal of Automobile Engineering*, 219, 135–149, 2005.
36. E. Ono, Theoretical understanding vehicle spin and stabilizing control, In *Proceedings of the Advanced Seminar on Vehicle Dynamics and Control*, JSME & JSAE No.06–16, pp. 45–66, 2006 (In Japanese).

33 Dynamics of Heavy Commercial Vehicles and Buses

John Aurell

CONTENTS

33.1 Introduction 1245
33.2 Historic Retrospect 1246
33.3 Vibration Environment 1247
33.4 Achievements 1248
33.5 Suspension Systems 1249
33.5.1 Chassis Suspension 1250
33.5.2 Cab Suspension 1253
33.5.3 Frame 1255
33.5.4 Engine Suspension 1255
33.5.5 Active Suspension 1256
33.6 Operating Conditions 1256
33.7 Handling 1258
33.8 Directional Response 1258
33.8.1 Tire Performance 1259
33.8.2 Suspension Effects 1262
33.8.3 Steering System 1263
33.8.4 Steered Rear Axles 1263
33.9 Stability 1264
33.10 Perception 1267
33.11 Vehicle Models 1268
33.12 Verification and Validation 1273
33.13 Conclusion 1276
References 1276

33.1 INTRODUCTION

The dynamics of heavy trucks, or heavy commercial vehicles as they are called in the ISO vocabulary, has obviously many similarities with the dynamics of passenger cars. There are however also some differences. This is due to the different areas and ways of use that call for different designs. While passenger cars are made to transport one or two or sometimes more people that often amount to less than 10% of the gross weight, trucks are made to transport as much goods as possible, with respect to weight or volume, within certain given weight and dimension limits [1].

One major difference is that heavy trucks, due to the laws of nature, are bound to roll over unless the driver is avoiding to exceed a certain lateral acceleration limit. Normally, passenger cars do not roll over, unless they are tripped, because the tires will saturate and start sliding long before

a sufficiently high lateral acceleration is reached. Exceptionally, however, it may occur in special maneuvers, like the so-called fishhook maneuver, with certain vehicle categories.

Another major difference is that heavy commercial vehicles and buses are quite often articulated with one or several articulation joints. This certainly makes the dynamics much more complex, as vehicle combinations have oscillatory modes that are excited in various maneuvers.

A third important difference is that, while passenger car chassis are fairly stiff with respect to torsion and bending, heavy commercial vehicles are normally not. This may have large effects on the dynamic behavior of the truck in various areas.

Although this chapter primarily deals with heavy trucks, much of the content also applies to heavy buses. Less research was carried out on heavy trucks and buses than on passenger cars, but there are still a large amount of papers dealing with various aspects of heavy vehicle dynamics [2] and also on more basic theories and surveys [3,4].

33.2 HISTORIC RETROSPECT

One could argue that much has not happened in the basic design of trucks since the middle of the last century. Trucks still have air-filled rubber tires, rigid axles, ladder frame with a cab on, some type of superstructure, and an internal combustion engine as on the Danish triple combination in Figure 33.1, with four articulation joints, from the late forties, transporting goods between Denmark and Germany.

This is true, but each of these systems has undergone a tremendous development, and the performance is so much better. There has been no revolution but rather an evolutionary process with continuous development. Figure 33.2 shows a Volvo Titan, which was still modern when the author



FIGURE 33.1 Old Danish triple combination.



FIGURE 33.2 Volvo Titan from 1954.

started his career at Volvo. It is a beautiful truck, belonging to Volvo's Veteran Truck Club, but it does not take a long ride to appreciate that the dynamic behavior is very far from a modern Volvo. It is definitely not as comfortable as it is beautiful.

The amount of transported goods has increased tremendously during the last half century. It has increased not only nationally but also, above all, internationally. This has resulted in that each vehicle is carrying more goods. The vehicles have become more specialized, and more weight optimized in order to be competitive. Also the engine performance has increased very much, giving higher average speeds. The demands on good ride for the driver as well as for the payload have been growing. So have the requirements on a high degree of availability of the truck. The traffic intensity is steadily increasing and this has put the active safety of the vehicles in focus. All these factors call for increasing vehicle dynamics considerations.

An interesting question is what factors are driving the development. The market is of course one. If there are no buyers of new developments, they will not take place. It is often said that the customer is always right, and that is true in one sense. It is however not entirely true, because the customer can only judge what already exists and is not always capable of assessing new features that may be very good. Another driver of the development is the company, the truck producer, and its core values. This is very important and assumes of course that the company is interested in driving the development in a certain direction and not only in making money. Finally, legislation and regulation may be an important driver, provided that the regulations are good regulations. If this is not the case, they may be counterproductive [5].

33.3 VIBRATION ENVIRONMENT

The vibration environment in a vehicle travelling on a given road at a given speed depends entirely on the dynamics of the vehicle. Good vibration environment means low dynamic forces and low acceleration levels. There are a number of aspects on the vibration environment, illustrated in Figure 33.3. The driver's vibration environment, the ride, is perhaps the most obvious, maybe because the driver can express what they perceive. It may also be the most important one, because human injuries, caused by vibrations, can rarely be repaired. Other aspects of vibration environment are related to the load, the vehicle structure, and the infrastructure [6–8]. The consequences of bad vibration environment for the driver may in the worst case be injuries from long-term exposure or fatigue decreased proficiency, detrimental to the traffic safety. The consequences for the cargo may be damages if the goods are sensitive. The infrastructure may suffer from increased wear and the vehicle structure from reduced fatigue life.

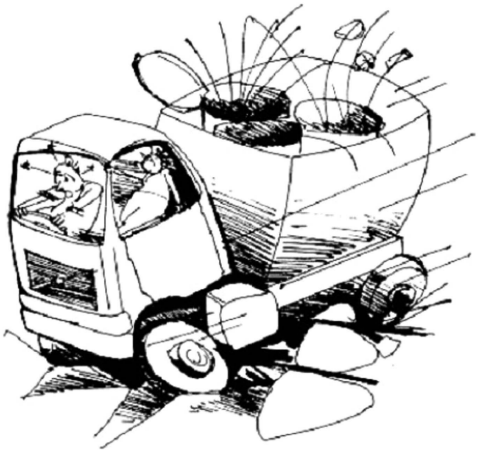


FIGURE 33.3 Consequences of bad vibration environment.

33.4 ACHIEVEMENTS

Large achievements have been made with respect to all four aspects of vibration environment mentioned earlier [9]. The most important is the improvement of the vibration environment for the driver, that is, in the cab. The reason for this is that if it is done in a clever way, the vibration environment for the rest of the vehicle, the load, and the infrastructure will also be improved as a spin-off effect. A clever way means that the dynamic forces shall be reduced as close as possible to their source, that is, in most cases the road. Another reason is of course the humanitarian aspect. There is absolutely no reason to accept that drivers should spend their working life in a miserable vibration environment. The current situation is not so bad, even if it varies throughout the world. The best trucks today offer a vibration environment close to that of a good passenger car.

A lot of research was carried out for many years concerning the effect of vibrations on the human body and how people perceive vibrations (see Chapter 36 of this book). It is a question of both amplitude and frequency. The human body is more sensitive in certain frequency ranges than in others. The reason for that is that the body may be seen as a biomechanical vibrating system. This is possible to model and the various body parts may then be seen as masses, springs, and dampers. The body has a number of resonances at which it has its highest sensitivity. For vertical whole-body vibrations, the maximum sensitivity is in the frequency range 4–12.5 Hz, and for horizontal vibrations, it is in the range 0.5–2 Hz. The way to measure, analyze, and assess whole-body vibrations is standardized, and the first edition of ISO 2631, appeared in 1974. It has since then been revised a couple of times. The evaluation is based on frequency-weighted root-mean-square (RMS) accelerations in three directions. The weighting filters have been discussed a lot and there have been many opinions concerning their shape. The present ISO filters are probably a good compromise, and since the sensitivity among individuals varies, there is no absolute truth [10]. Figure 33.4 shows asymptotic approximations of the weighting curves. *Wd* is the curve for horizontal accelerations and *Wk* the curve for vertical accelerations.

There is a growing attention from society on the potential risk with vibration exposure. Since 2002, there is a directive in the European Union concerning limits for whole-body and hand-arm vibrations, and it is now being implemented in national regulations [11]. For whole-body vibrations, there are two limits for 8 h exposure time. One is an action value of 0.5 m/s^2 and the other is an exposure limit of 1.15 m/s^2. These values are weighted RMS accelerations, in accordance with ISO 2631, measured over the whole 8 h working day. The important shortcoming with only using wideband RMS values is that these assume more or less stationary vibrations and do not take

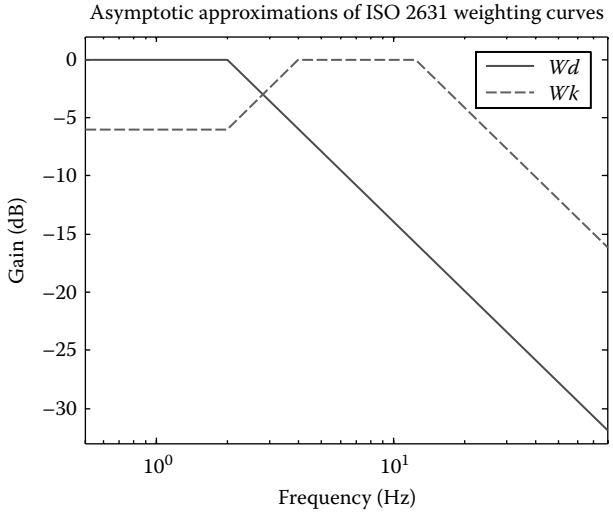


FIGURE 33.4 Weighting curves for horizontal (*Wd*) and vertical (*Wk*) accelerations.

shocks and transients into account. This new directive will anyhow put the vibration environment in focus and will probably drive the development.

The effect of heavy vehicles on wear and damage of the infrastructure has been studied in many large research projects lately. It is well known that the static axle load has a major influence and much of the research was focused on the effect of dynamic axle loads. This resulted in a definition of the so-called road-friendly suspensions, friendly in the sense that they provide low dynamic wheel forces. To fulfill the requirements for road friendliness, the dominant eigenfrequency, obtained when the axle is excited with a step input, must not exceed 2 Hz, and the relative damping must be at least 20%. In practice, this demand is difficult to reach without air suspension. In order to promote road-friendly suspensions, existing regulations use incentives like increased gross weight or axle load. Although the definitions may be discussed, it is basically a good way of using regulations [12–15].

33.5 SUSPENSION SYSTEMS

The key elements for a good vibration environment are well-functioning suspension systems. In heavy trucks, there are several important systems. Starting from the ground, there are the tires, the chassis suspension front and rear, the frame, the engine suspension, the cab suspension, and sometimes a seat suspension. They are all there to provide vibration isolation. It is important that they are matched together from the start. If not, the risk for suboptimization is obvious. Figure 33.5 illustrates the principal isolating effect of the various suspension systems.

The lower the rigid body eigenfrequencies are, the better the potential for good vibration isolation is. This is because the road input is basically broadband random excitation. There are three basic characteristics of suspensions: stiffness or compliance, damping, and space or spring travel. The lower the stiffness, the more compliant the suspension is, and the better the vibration isolation. In other words, the softer the suspension, the better it is. But in order to be able to make a soft suspension, there must be enough space. This is a law of nature and may not be violated without severe punishment. The punishment consists of transients, due to bottoming, when the road gets rougher. The third characteristic is damping. It is necessary to dissipate energy and limit the amplitudes at resonance frequencies. There is an optimum value of the damping that gives minimum vibration response. It is also important to minimize low-frequent movement both vertically and in particular in roll in order to get a feeling of good control. By using nonlinear damper characteristics in a clever way, it is possible to damp these slow, large movements without interfering too much with the vibration isolation, that is, to achieve good control without harshness and bumpiness. Principally it is better to have a low spring rate with high damping than a higher spring rate with less damping (see Chapter 8).

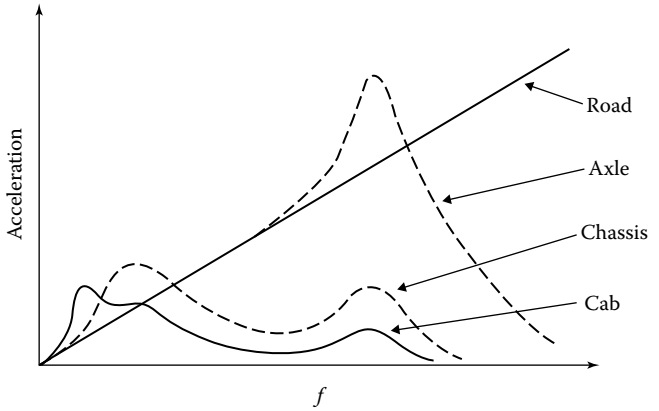


FIGURE 33.5 Vibration isolation effects of various suspension systems.

33.5.1 Chassis Suspension

Damping shall ideally be viscous, but some spring elements, for example, leaf springs, have more or less inherent friction, which also provides damping due to the hysteresis. The friction does however not only provide damping but also stiffness, especially at small amplitudes. Figure 33.6 shows the behavior of a leaf spring. It can be modeled with a model developed at University of Michigan Transport Research Institute (UMTRI) that has an upper and a lower stiffness envelope and an exponential approach to the envelopes. The rate of the approach is described by a special parameter [16].

The friction causes a phase shift between force and displacement, which introduces equivalent viscous damping. The equivalent viscous damping and stiffness of a leaf spring with coulomb friction may be estimated from experiments, where the spring is excited with realistically shaped noise. The results are obtained in the frequency domain. Figures 33.7 and 33.8 show damping and stiffness for three excitation levels as a function of the excitation frequency. It appears that both damping and stiffness are higher for low excitation levels. This implies that the leaf spring is acting as a much stiffer spring on good and medium roads. The nominal spring rate is obtained only on very rough

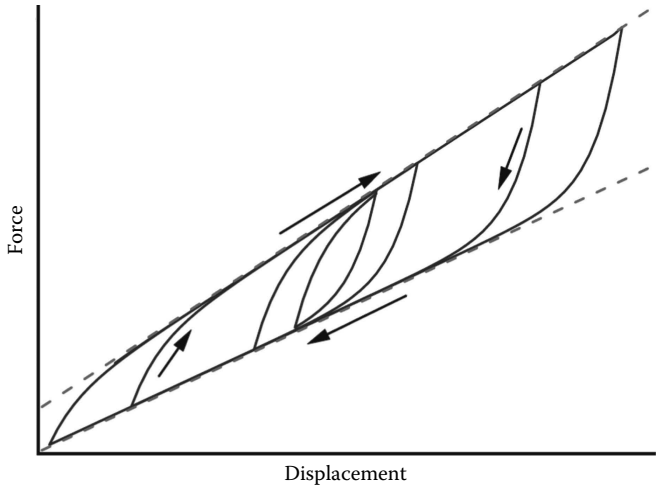


FIGURE 33.6 Characteristics of a leaf spring.

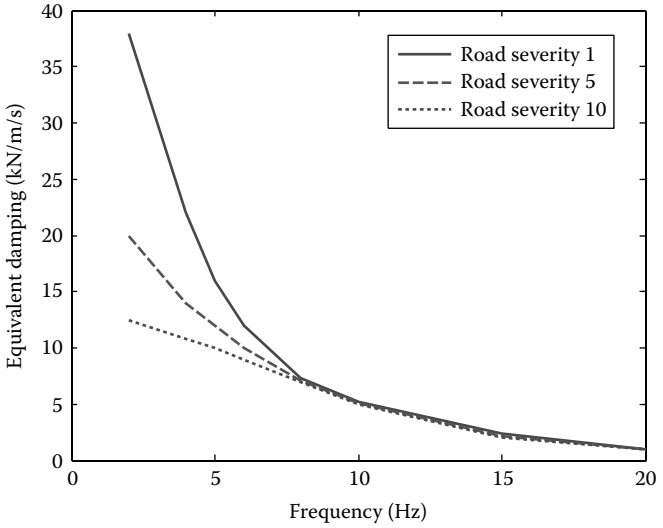


FIGURE 33.7 Equivalent damping of a leaf spring.

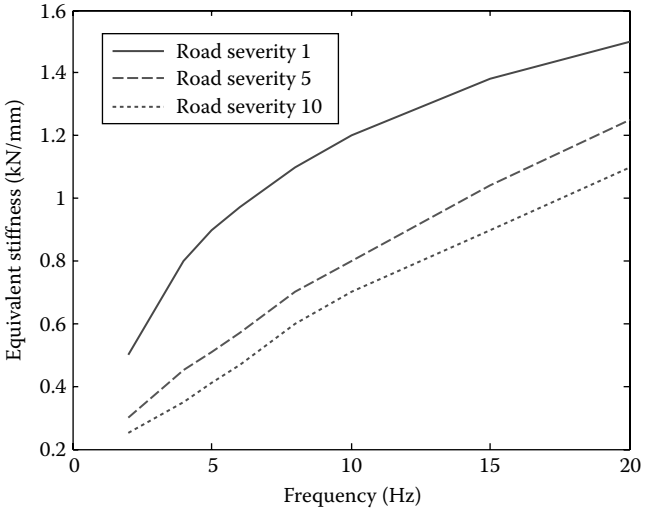


FIGURE 33.8 Equivalent stiffness of a leaf spring.

roads at low frequencies. While the damping is decreasing with frequency, the stiffness is increasing. The reason for this is that the leaf spring may be regarded as having stiffness not only in parallel with a viscous damper but also in series (see the Cab Suspension Section 33.5.2).

The friction impairs the ride significantly on good roads, but on rough roads, it affects less. One positive effect is however that the low-frequent movements are more efficiently damped, which improves both the vertical and the roll control. It is thus important to avoid and reduce friction in suspension systems. For leaf springs, this may be achieved by putting rubber spacers between the leaves. This is of course only possible on parabolic springs, as the one shown in Figure 33.9.

One characteristic of heavy trucks is the large load variation between laden and unladen condition, especially on the rear axles. This is a limiting factor for low spring rates of mechanical suspensions without level control. On front axles, however, the difference is rather small, and it is possible to use soft suspensions without leveling systems. This is one reason for still using leaf spring suspensions on the front axle of high-performance trucks. With high-quality steel, it is possible to make leaf springs with sufficiently low stiffness with few leaves. The spring rate range expressed as eigenfrequency of a quarter car model is thus different in front and rear for mechanical suspensions. Front suspensions are in the range 1.3–1.8 Hz, while rear suspensions have frequencies in the range 2–3 Hz.

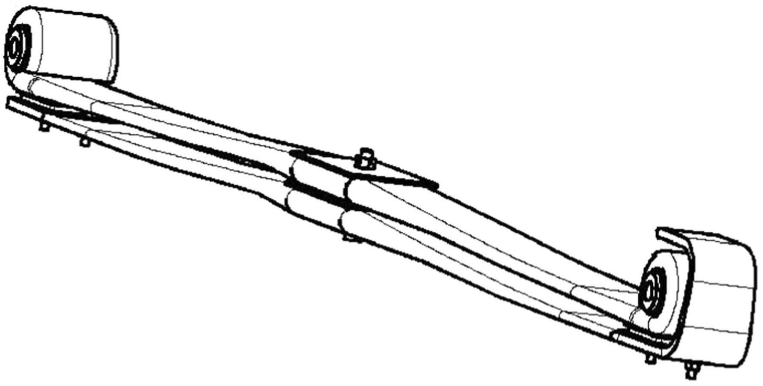


FIGURE 33.9 Parabolic leaf spring with rubber spacers.

An axle suspension shall provide both vertical compliance and roll stiffness. If these two features are not separated, but provided by the same spring element, it will usually end up in a poor compromise. The vertical stiffness will be too high and the roll stiffness too low.

Air suspensions were introduced in the 1980s and have since then gradually taken over most of the market on long haul trucks, at least on rear axles. There are a number of advantages with air suspensions. One is the built-in leveling, which keeps the vehicle at a constant ride height and facilitates load handling, for example, of swap bodies. Another big advantage is that air suspensions easily provide low spring rates, and due to the leveling system, it is possible to have low spring rates. A third one is that air springs with constant air volume provide spring rates proportional to the load, which gives a constant eigenfrequency and as a consequence good vibration isolation also in partly laden conditions. A fourth one is that the friction and the parasitic stiffness are normally very low in air suspensions. The negative aspect is that air suspensions are more expensive than leaf spring suspensions. The eigenfrequency range for air suspensions is 1–1.5 Hz.

Figure 33.10 shows a Volvo 6 × 4 rear air suspension, also available as 6 × 2 and 4 × 2. The axle is controlled by two longitudinal reaction rods and one V-stay. The axles can move vertically and in roll. The absence of pitch rotation prevents driveline vibrations. Vertical and roll stiffness are separated. Four air bellows provide the vertical stiffness, while the roll stiffness is obtained by a stabilizer.

Damping of the suspension is very important. Old trucks relied entirely on friction damping, but to achieve a good vibration environment, the damping should be viscous. Ideally the dampers should only damp the resonances, but as the damp forces depend on the relative velocity and not the frequency, this is not possible. Therefore, the dampers do not only damp the resonances but also increase the force transmission for frequencies between resonances. It is therefore necessary to optimize, tune, the damp characteristics with respect to various demands on vibration environment and handling. Modern passive dampers have good possibilities to adjust the forces. It is usually possible to shape the damp characteristic within three velocity ranges. Figure 33.11 shows an example of a damper characteristic. The compression forces are normally much lower than the expansion forces in order to reduce the force transmission into the vehicle. Too low compression forces on the other hand reduce the control or "stability" of the vehicle. The forces at very low velocities, below 0.05 m/s, are important for the control, both in roll and vertically, and a high damping rate is advantageous. Damping characteristics are usually very degressive in order to reduce bumpiness. The forces at high and medium velocities shall not be higher than is required to damp axle resonances.

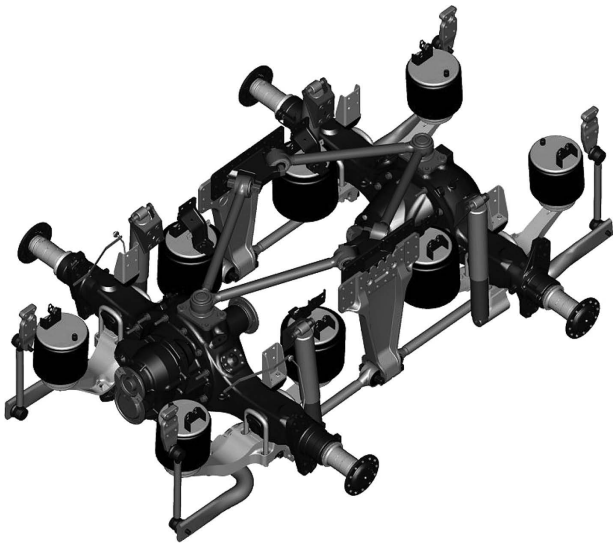


FIGURE 33.10 6 × 4 rear air suspension with reaction rods and V-stays.

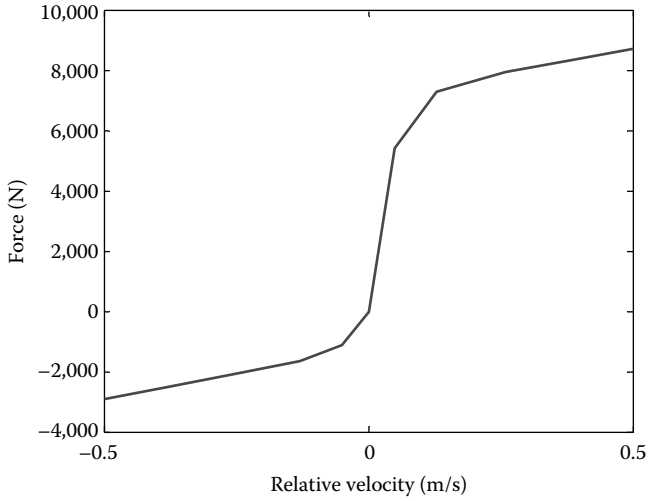


FIGURE 33.11 Example of damper characteristic.

The majority of heavy trucks are equipped with rigid axles. The majority of heavy buses are however equipped with independent front wheel suspension. Due to the large benefits for most applications, we will see a development with independent front wheel suspension also on heavy trucks. The main advantages are improved vibration environment and handling. Rear axle suspensions are however likely to keep the rigid axles for most applications.

33.5.2 Cab Suspension

The first fully suspended cabs appeared in the 1970s. Before that, the cabs were mounted either on stiff rubber mounts in four points or on two stiff mounts in the front and two softer springs in the rear. These arrangements gave very poor vibration isolation or even amplification. The very close relationship between vibration environment and fatigue life was illustrated very clearly in the development of Volvo F10/12, which was launched in 1977. Comparative accelerated endurance tests were then run between two vehicles, where one was equipped with a four-point soft cab suspension and the other with a traditional one with two stiff and two soft suspension points. Figure 33.12 shows the number of faults vs. cycles of the endurance test on the proving ground. The outcome of these tests was very convincing. The number of faults on the fully suspended cab was less than half of the number of faults on the other one.

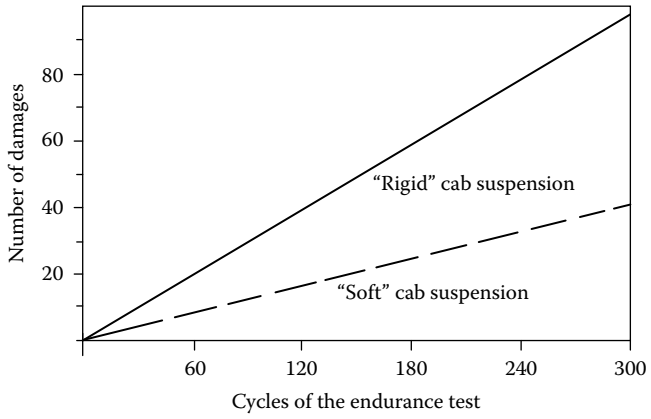


FIGURE 33.12 Comparison between soft and rigid cab suspension.

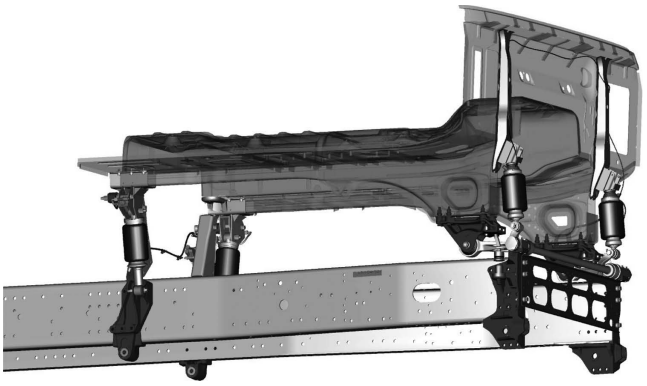


FIGURE 33.13 Example of a modern cab suspension.

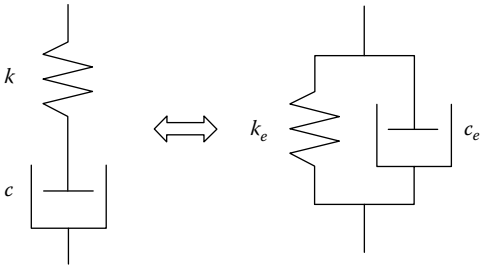


FIGURE 33.14 Stiffness in series converted to stiffness in parallel.

Since then, the use of air-suspended cabs has increased, and today air springs are predominantly used, for example, as the cab suspension shown in Figure 33.13. The basic principle is to make the bounce and pitch eigenfrequencies low and uncoupled in order to be able to isolate especially the pitch of tractors. The roll stiffness must be sufficiently high and is provided by an anti roll bar. It is obviously necessary to tune the cab suspension with the chassis suspension. The eigenfrequencies of the cab suspension should not coincide with other resonances. Lateral and longitudinal stiffness must be carefully considered together with the complex structural behavior of the front end of the chassis and with the engine suspension.

The design of the attachment points between the cab and the frame should also consider the transmission of structure-born sound. This is an area of conflict, and compromises have to be found. One example is the damper attachment. From a noise transmission point of view, this attachment should have a low stiffness, because noise is transmitted through the damper. From a vibration isolation point of view, it should be as stiff as possible. The reason for that is that a spring in series with a damper reduces the damping and adds stiffness in parallel with the damper as shown in Figure 33.14.

The following expressions for k_e and c_e, as a function of frequency f, can be derived:

$$k_e = \frac{k}{(1+(k/(2\cdot\pi\cdot f\cdot c))^2)}$$

$$c_e = \frac{c}{(1+((2\cdot\pi\cdot f\cdot c)/k)^2)}$$

The equivalent stiffness can be quite substantial in comparison with a soft cab spring. This is illustrated in Figure 33.15, which shows equivalent, frequency-dependent, stiffness, k_e, and damping, c_e, for a damper in series with a spring for the case when the damping rate c = 10,000 Ns/m and the spring rate k = 1,000 N/mm.

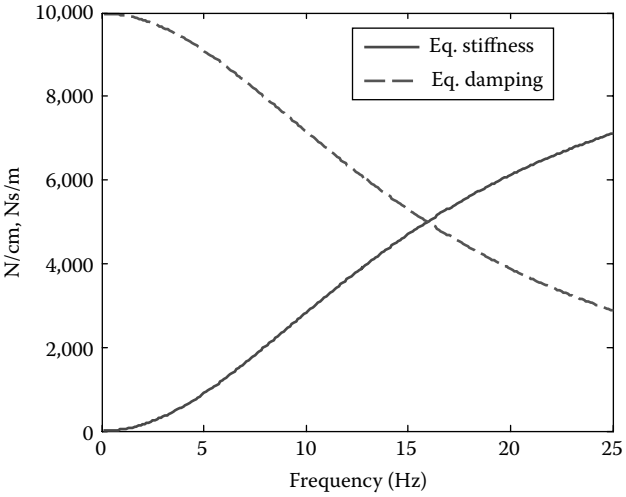


FIGURE 33.15 Damping and stiffness of a damper with elastic attachments.

It is also important, though difficult, to design the cab suspension in such a way that friction and parasitic stiffness are minimized. The cab suspension must on the other hand be well damped with viscous dampers. The relative damping is normally much higher than for the chassis suspension and often close to critical damping.

33.5.3 Frame

The frame is not really a suspension system, but it has a special impact on the vibration environment due to its flexibility. There are a number of internal excitation sources. One of these is tire runout, which is a periodic excitation. This may excite a phenomenon called "jumping ride" or "beaming" that occurs on rigid trucks and on tractors with exceptionally long wheelbase. It is a frame vibration and the frame vibrates with its first bending mode. The frame vibration causes a pitch movement of the cab, which is felt by the driver as a longitudinal vibration. It usually occurs in the frequency range 6–8 Hz and corresponds to the speed range 65–85 km/h. It is normally a narrow and undamped resonance peak. Although the amplitude is small, the vibration is very annoying for the driver. Weight optimization of the trucks easily worsens the situation, because that may imply reduced stiffness of the frame. The peak of this phenomenon was in the late 1960s and in the 1970s. An example, from this time, shows the vibration modes for the first two bending modes and the bounce and pitch modes in Figure 33.16. The roads had at the time become quite good. It sounds like a paradox, but even roads are a prerequisite; otherwise, this resonance is concealed in the noise. There were basically two ways to cure this problem. One was to make the tires round, and the other was to stiffen the frame with reinforcements. Neither solution is ideal. Another more sophisticated method was also successfully used. That was to use the engine as a dynamic damper by tuning the engine suspension. It requires a certain geometry of the engine suspension to be able to do that. The final solution to this phenomenon appeared when fully suspended cabs were introduced. The frame vibrations are still there, but they are effectively isolated from the cab.

33.5.4 Engine Suspension

The engine is an excitation source. The task of the engine suspension is to isolate the engine vibrations from the chassis. The engine has many excitation orders, but the most important for a six-cylinder engine is the third order. The highest eigenfrequency of the engine suspension must be well below this excitation frequency at idling, that is, ca. 25 Hz. The engine has a substantial mass and inertia that is

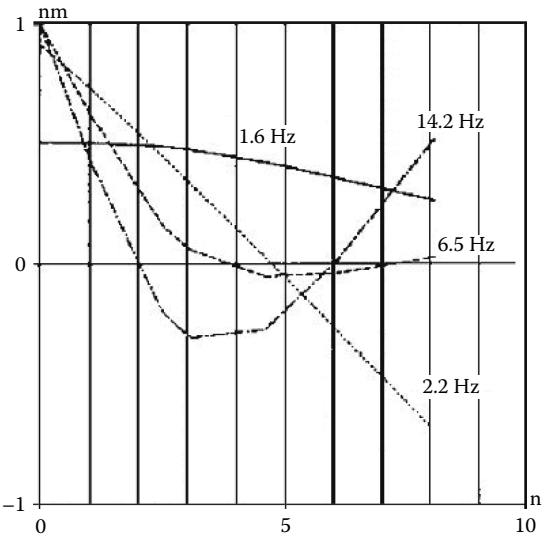


FIGURE 33.16 Mode shapes of a frame.

of at least the same magnitude as the other large masses in the front end, the cab, and the front axle installation. It is therefore important that the engine suspension is tuned in order to avoid uncontrolled dynamic couplings. The range of the eigenfrequencies of the engine suspension is 7–15 Hz.

33.5.5 Active Suspension

In fully active suspensions, spring and damper elements are replaced by a force generator, which is controlled by state variables of the vehicle (see Chapter 31). This implies that there is both energy input and energy dissipation in the suspension. The active suspension may be optimized with respect to a weighted sum of body acceleration (ride), tire force (road holding), and suspension travel (space). Another application is active stabilizers. Very good frequency response of the complete system is required in order to avoid high-frequent harshness. There are two important disadvantages, high cost and high power consumption. This makes fully active suspensions not suitable for trucks at the present stage. Semiactive suspensions are from these points of view more attractive. They have an actively controlled damper and can only dissipate energy. The principle of the control is the so-called skyhook damping, in which the damper is imagined to have one end fixed in the sky and thus being able to damp the absolute velocities of the body and the wheel instead of the relative velocity between the body and the wheel. Also here, the frequency response must be very good to avoid harshness. The ride improvement with well-functioning semiactive chassis suspensions is 10%–15%, measured as RMS acceleration. This is actually not enough to justify such a complex and expensive system. The fact is that passive suspensions have a very large potential. By using soft chassis springs and well-optimized, nonlinear dampers, very good results can be achieved. There is however a third possibility, and that is adaptive dampers. By adapting the damping to the load, to the road input, or to roll excitation, better performance may be achieved in certain situations. Another potential possibility would be to adapt the dampers to reduced space for the suspension in order to avoid bottoming. Adaptive damping is also a simpler system but still much more expensive than passive dampers.

33.6 OPERATING CONDITIONS

There are a number of internal periodic excitation sources in the engine, transmission, and wheels that are important for the ride comfort. The main excitation source is, however, the road unevenness. This is also the only source of large dynamic forces that are important not only for the ride comfort

but also of interest for fatigue and damage of the vehicle structure and wear of the infrastructure. As the nature of the road excitation is stochastic, the longitudinal road profile may be described by an autospectrum in the spatial frequency domain [17,8]. The road profile may also be assumed to be isotropic, that is, it has the same road spectrum in all directions. The nature of the road excitation was first described on runways of airfields, and since then, a large number of road profiles have been measured and analyzed. The results all show that a straight line in a log–log diagram, that is, with the following equation, may represent a road spectrum:

$$\Phi = \Phi_0 \cdot \Omega^{-w}$$

where

Φ is the power spectral density of the road profile
Ω is the spatial angular frequency
w is the waviness
Φ_0 is the road severity, the power spectral density at $\Omega = 1$

For a long time, a waviness of 2 was commonly assumed and used in simulations and elsewhere. The reason for that may be that it is simple, but it is not correct. For the most common road network, the waviness is rather 3, that is, the content of long wavelengths is larger. For networks of bad quality, the waviness may be 2.5, and only for extremely bad roads, the waviness is as low as 2. Figure 33.17 shows an average road spectrum for 25 km of a bad minor road and a linear curve fit.

Another common mistake that was made is the assumption that roads are stationary with a constant road severity. But the fact is that they are not. This may easily be overlooked because the RMS values for various responses would be the same for reasonably linear trucks for both stationary and nonstationary input. However, if there are nonlinearities when fatigue phenomena of the vehicle structure or of the human body are to be investigated, it will give completely different results whether nonstationary inputs are considered or not. In order to be able to optimize and differentiate the trucks for various applications and markets, it is necessary to describe the nonstationary excitation.

One way to describe the nonstationary road input is to use the Q-distribution. Q is a combined parameter of road severity, vehicle speed, and waviness and may be used to define the operating conditions [18]:

$$Q = \Phi_0 \cdot v^{w-1}$$

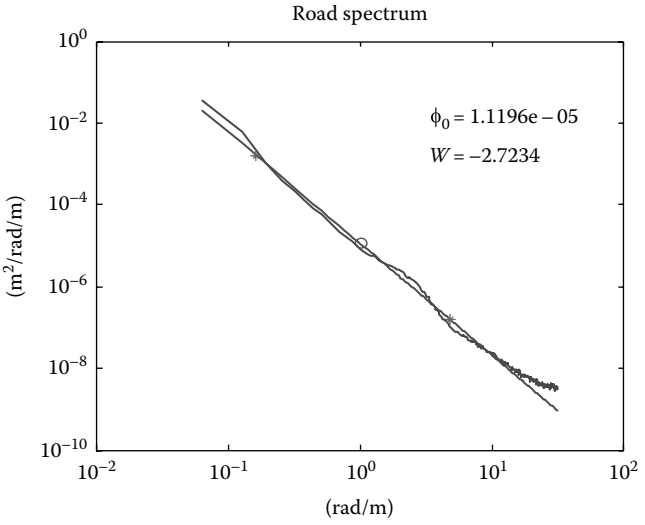


FIGURE 33.17 Measured road spectrum.

The Q-distribution is expressed by a Weibull distribution, which has the probability-density function:

$$f(Q) = \frac{\beta}{\theta}\left(\frac{Q}{\theta}\right)^{\beta-1} e^{-(Q/\theta)^{\beta}}$$

where different operating conditions are defined by different sets of Weibull parameters β and θ.

Another way is to define a discrete number of road severities based on the Q-distribution and then for each road severity describe a distribution of transient inputs.

The description of the nonstationary excitation is based on extensive measurements in real operations [19].

33.7 HANDLING

Good handling of a truck is driven by active safety requirements. Handling and stability are the most important features for active safety. Transport effectiveness is another important driver for handling and stability features. Vehicle combinations are configured in order to optimize the weight or the volume of the payload. It is vital that this is done without degrading the stability. Higher speeds also increase the requirements on the handling and stability of heavy trucks. The driving pleasure, finally, is also an important factor.

The handling feature may be divided into three areas. One is the directional response of the vehicle or the vehicle combination due to a steering input or to a road input. The second response of importance is the roll behavior due to steer input or road input. The third area of interest is the interface between the driver and the vehicle, that is, the driver's perception of the vehicle and its controls. The two first areas deal with objective handling characteristics and the last one with subjective characteristics.

33.8 DIRECTIONAL RESPONSE

The steady-state behavior of a truck is determined in a circular test procedure, where either the radius or the longitudinal velocity is constant [43]. The relationship between steering-wheel angle and lateral acceleration is the most basic one. The derivative of the steering-wheel angle, divided by the steer ratio, with respect to lateral acceleration gives the understeer gradient, which is an important characteristic value. The understeer gradient shall be positive. A negative understeer gradient gives instability at a certain critical velocity. The relationship between steer angle and lateral acceleration should preferably be linear, but if it is not, the understeer gradient should increase with lateral acceleration. For heavy vehicles with more than two axles, the relationship between steer angle and lateral acceleration is speed dependent, which is not the case for simple vehicles with two axles. By using the concept equivalent wheelbase, l_e, multiaxle vehicles may be treated as simple vehicles [20]. Figure 33.18 shows a handling diagram in a constant velocity test with the three-axle rigid truck in Figure 33.19 [21]. The different behaviour in left and right turns is due to assymetric stiffness of the steering system.

The understeer gradient is obtained from $(\partial\delta_H/\partial a_c)\times(1/i_s) - (\partial\delta_D/\partial a_C)$ where $\delta_D = l_e/R$.

The transient behavior of a truck may be determined in a number of different test procedures [22]. Most often the yaw velocity and the lateral acceleration responses in relation to the steering wheel input are studied. Since the lateral acceleration response is dependent on where it is measured, the yaw velocity response is preferably used. If the truck behaves linearly, which it often does, methods employing analyses in the frequency domain may be used. Analyses in the time domain usually yield results only for specific maneuvers. One important characteristic value is a quantification of the time delay between yaw velocity and steering-wheel angle, such as response time, time lag, phase time, or phase angle. Another characteristic value is the yaw velocity to steer angle gain. A third important characteristic value is the frequency bandwidth of the yaw response.

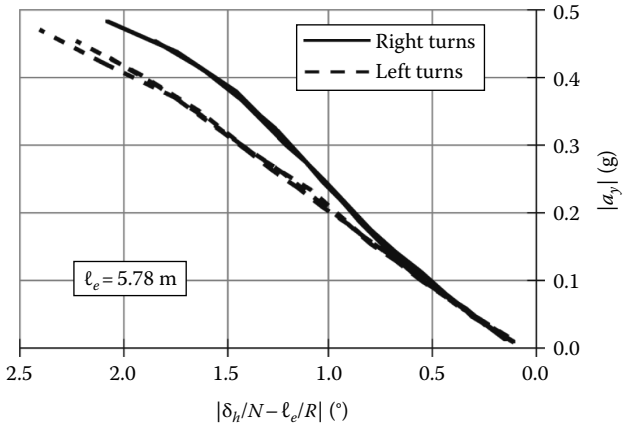


FIGURE 33.18 Handling diagram (Note: δ_h is the steering-wheel angle, $a_y \approx a_c$ is the lateral acceleration, $N = i_s$ is the steer ratio, R is the radius).



FIGURE 33.19 6 × 2 rigid truck.

33.8.1 Tire Performance

The magnitude of the forces between wheels and ground and how they are generated determine the behavior of the truck in all situations (Figure 33.20).

Tires are thus the single most important component of the truck for its handling characteristics and stability. The relationship between lateral, cornering, tire force, and tire slip angle is very non-linear, seen over a large span of the slip angle. Figure 33.21 shows an example of the relationship between lateral force and slip angle for various vertical forces. These curves are based on the "magic formula" of which the early development took place at Volvo in cooperation with Delft University of Technology and now is the widely accepted way of presenting tire data (see Chapter 17).

As trucks however very seldom are driven close to the limit of the tires, that is, at large slip angles, but rather at slip angles normally not exceeding 2°, the relationship may be considered as a linear one and can be expressed with the cornering stiffness, C_α:

$$c_\alpha = \left(\frac{\partial F_Y}{\partial \alpha} \right)_{\alpha=0}$$



FIGURE 33.20 Truck in equilibrium.

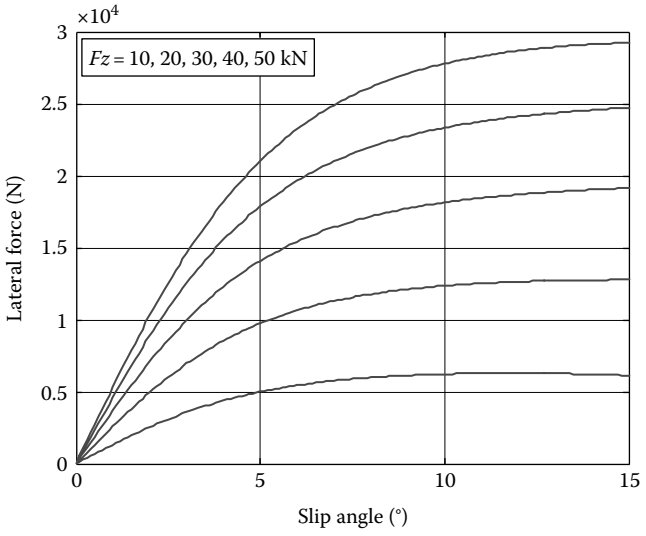


FIGURE 33.21 Example of tire characteristics.

There is still one nonlinearity of importance, and that is the relationship between cornering stiffness and the vertical tire force. Figure 33.22 shows this relationship for the same tire as previously mentioned.

This nonlinearity causes a reduction of the total axle cornering stiffness due to load transfer between left and right wheels. This reduction is more pronounced when the load transfer is large, that is, for high lateral accelerations.

It is advantageous to use cornering compliance instead of cornering stiffness. The cornering compliance is the inverse of the cornering coefficient, which is the cornering stiffness normalized with the vertical force. The difference between cornering compliance front and rear determines the understeer/oversteer behavior. It is therefore extremely important to have the correct tire combination front and rear. The understeer gradient may vary with the lateral acceleration, mainly due to tire nonlinearities. Obviously, also the load distribution and the center of gravity height affect the understeer gradient.

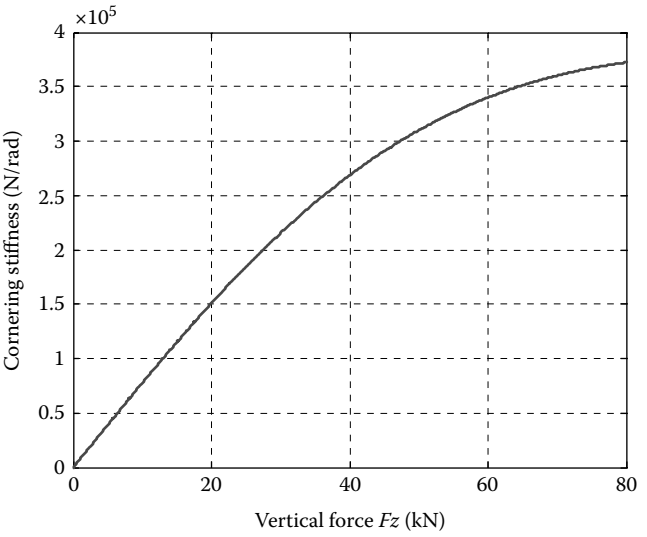


FIGURE 33.22 Nonlinear cornering stiffness.

There are large differences between passenger car tires and truck tires. Good truck tires have a cornering compliance around 8°/g, while good passenger car tires have not more than a fourth, that is, around 2°/g. This depends on that tire life and rolling resistance are very important for transport economy and are prioritized for truck tires. This is the main reason for the quicker response of passenger cars.

The state of wear has a large effect on the cornering characteristics of truck tires because a large part of the compliance is located to the thread. The cornering stiffness may increase between 25% and 50% when tires wear. Figures 33.23 and 33.24 show an example on the effect of worn tires on the frequency response of yaw velocity for a 4 × 2 tractor/semitrailer combination [23].

It appears that the bandwidth increases significantly and the phase delay is reduced with the worn tires.

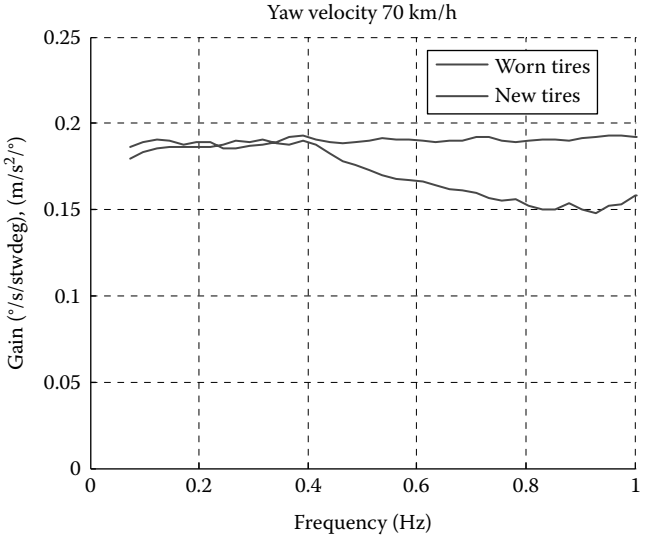


FIGURE 33.23 Effect of the state of wear on gain.

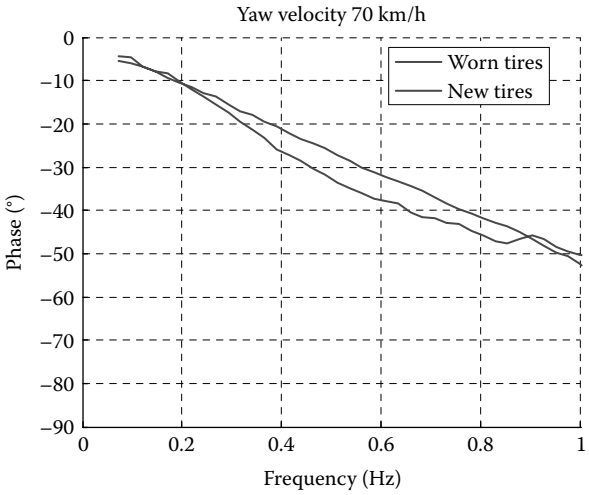


FIGURE 33.24 Effect of the state of wear on phase.

Tire characteristics, obtained from measurements, are dependent not only on the state of wear but also on the temperature and on the measuring equipment. Comparisons of measurements therefore have to be done with caution.

There has been a tremendous development of truck tires with respect to most characteristics such as cornering capability, life, and rolling resistance. One major step was the shift from cross-ply tires to radial tires. Low-profile tires were developed for passenger cars in order to reduce the cornering compliance. For trucks, low-profile tires are used primarily to reduce the outer diameter of the tires in order to be able to increase the load volume.

Another difference between passenger car tires and truck tires is the change of the cornering stiffness when the tire pressure is reduced. While the cornering stiffness is often reduced for passenger car tires, it increases for truck tires. There is obviously a limit for this increase. If a tire explosion occurs and the tire is completely deflated, the cornering stiffness will also be close to zero. This obviously changes the characteristics of the vehicle drastically and may cause instability. If dual tires on the rear axles are replaced by singles, special precautions must be taken to prevent the risk for instability. Tire explosions on front axles may however also cause accidents. The side pulling in such a situation is moderate and the truck is easy to steer and brake, but if the driver does not have full attention on the driving for some reason, they or she may be reacting too slowly and an accident may happen. The conclusion of this is that there is a need for run-flat tires on trucks, which can maintain most of their characteristics at a sudden pressure loss.

33.8.2 Suspension Effects

Induced steering effects by the suspension, such as roll steer and compliance steer, create a cornering compliance of the suspension in terms of steer angle/acceleration. This adds to the tire cornering compliance and contributes to the total cornering compliance for each axle. Thereby, both the understeer/oversteer behavior and the transient response are affected. It is advantageous to let the induced steering at the rear axles be induced understeer, which means that the cornering compliance of the rear is reduced. This has a similar effect as to increase the cornering stiffness of the rear tires. Compliance-steer effects are better than roll-steer effects because they are in phase with the lateral force buildup, while the roll-steer effects are delayed. Induced understeer at the front axle also increases the understeer gradient. This is not as advantageous as at the rear axles because it will not improve the transient characteristics as effectively.

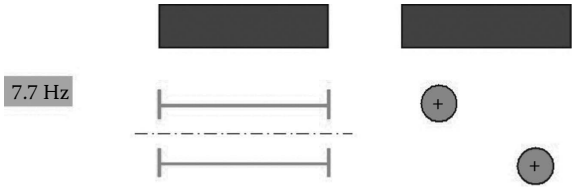


FIGURE 33.25 Front and side view of a bogie pitch mode.

Poorly damped suspensions may lead to loss of road grip when the wheels are bouncing at axle resonance. This leads to reduced lateral force capability of the tires due to relaxation effects. Loss of road grip at the rear axles leads to instability and at the front axle to an unsteerable vehicle.

Bogies of walking-beam type can easily loose the road grip, especially when unladen or partly laden, because they have an additional, very lightly damped mode, a bogie pitch (see Figure 33.25). On rough roads, this may lead to a bogie skip, which may cause a jackknife. It is therefore important that bogies are equipped with viscous dampers.

33.8.3 Steering System

The steering system is for many reasons important for the handling characteristics. Flexibility between the steering wheel and the road wheels adds to the total cornering compliance of the front axle and increases the understeer of the vehicle. Without this flexibility, many trucks would actually be oversteered. Front suspensions of heavy trucks most commonly use leaf springs as spring element, due to their cost-effectiveness. It is likely to stay this way as long as rigid axles are used. This makes the design of the steering geometry more difficult. In order to control steering errors, the kinematics of the leaf spring must be well known, for example, position of Ross center and inclination of Ross line. Sometimes steering errors are deliberately built in to give roll steer in order to increase the understeer. This is not recommendable, because steering errors may also cause both brake steer and ride steer at unacceptable levels. To achieve good steering with minimized errors in all respects, not only the kinematics of the steering but also all flexibilities have to be considered.

Another phenomenon is steering kicks, which are characterized by high transient forces in the steering linkage that are transmitted to the steering wheel. They may occur in potholes and on bad roads and are caused by steering errors and unfavorable kinematics of the front suspension. In such a situation, it is however easy to jump to conclusions, because the steering kicks occur at axle resonance. A substantial increase of the axle damping eases the problem, but it may then cause large ride problems. Another not-so-clever way of reducing steering kicks is to increase the power assistance in the steering gear. The downside of that is that the steer feel becomes impaired.

Friction in the steering system is unavoidable, and it is both a good thing and a bad thing. It is good because damping is necessary in the steering system in order to avoid shimmy and wobbling. A sufficient amount of friction makes it possible to avoid a viscous damper in the system. It is bad because it deteriorates the steer feel. A good compromise is necessary. However, friction between the steering wheel and the steering gear should be minimized.

33.8.4 Steered Rear Axles

In order to improve low-speed maneuverability and to avoid wheel scrub, heavy trucks use more and more frequently steered rear axles. This affects the directional response. There are two types of rear-wheel steer. The most simple one is self-steering, which implies that virtually no lateral tire forces are generated. Self-steering will result in shimmy unless there is a high amount of friction in the steering. This happens because the lateral tire forces will lag behind the steer angle due to tire relaxation. Consequently, self-steered axles require dampers in order to avoid shimmy. The other one is forced or positive steering, which means that the steer angle of the rear wheels is proportional

to the front steer angle. In this case, lateral tire forces are generated but at a lower level than in the case with fixed axles. In both cases, the steering results in increased cornering compliance at the rear, which gives less understeer and a slower transient response [24]. Because of this impairment of the handling, the steered axles should be locked above a certain, rather low speed. For the self-steer case, this must always be the case. The same thing applies to trailer applications. The forced-steer case gives obviously a potential to steer the rear wheel in such a way that the response of the truck becomes quicker at normal speeds, if the rear wheels are steered "by wire." This may be done by simply countersteering the rear wheels proportionally to the front wheels or to steer them actively, controlled by yaw velocity or other state variables [25].

33.9 STABILITY

Yaw instability of a single-unit truck rarely occurs. It requires an oversteered vehicle driven above its critical speed or special maneuvers. It is more likely to happen on low—than high—friction surfaces. Special circumstances, for example, an empty tractor that is braked during cornering, may however lead to aperiodic yaw instability. Roll instability, on the other hand, happens frequently. It occurs easily in a steady-state or quasi steady-state situation, when the rollover threshold is exceeded. There is a very strong correlation between the steady-state rollover threshold and the risk for rollover in real traffic [26]. The rollover limit is mainly determined by the track and the center of gravity height, but it is also influenced by other factors such as various sources of compliance and lash. The suspension compliance and lash are important but also the distribution of roll stiffness between the axles. The torsional compliance of the frame can have quite an impact as is illustrated in Figure 33.26. Turning the skilifts 180° shifts the weight closer to the axles and reduces the deflection of the frame substantially, which increases the rollover threshold.

Single-unit vehicles are mainly used as construction or distribution trucks. In long haul operations, they are very rare, and the reason is of course transport effectiveness. It is a matter of getting as large a volume as possible within given constraints.

As vehicle combinations are much more complex than single-unit vehicles, one may run into problems, if the dynamics of various types of vehicle combinations is unknown. The specific behavior of vehicle combinations in general is that the response to a steering input is an oscillatory, more or less damped, movement. These oscillations may in the worst case lead to a true instability where the damping is negative and the oscillations become self-excited. In other cases, they may just lead to an amplification of the movements of the first vehicle unit, the motor vehicle, to the last trailer



Rollover threshold = 0.24 g;	Bucket to bucket gives 0.31 g

FIGURE 33.26 Rollover test at UMTRI.



FIGURE 33.27 Vehicle combination with stability problems.

in the vehicle combination. A high rearward amplification of the lateral acceleration may lead to overturning of the last trailer. High rearward amplification of the yaw velocity may lead to that the last trailer is sliding off the road or hitting other vehicles. Commonly used characteristic values for dynamic stability are rearward amplification and damping. Figure 33.27 shows a car transporter consisting of a truck/central-axle trailer combination. This particular combination appeared to be weakly damped and very sensitive. Oscillations were easily excited by steer inputs, road ruts, and wind gusts.

Each type of vehicle combination has to be investigated separately, but there are a couple of guidelines, generally valid for maximum dynamic stability. Some of these are that the number of articulation joints shall be as small as possible, the overhang to coupling points shall be minimized, the wheelbase of trailers shall be long, the trailer tires shall have high cornering stiffness, and of course the height of the center of gravity of all vehicle units shall be as low as possible [27–29].

There is a strong trend toward high cube goods. This results in low chassis, since downward is the only way that the volume can be increased. The inevitable consequence of this is tires with smaller diameter and suspensions with shorter spring travel, and this is detrimental both to the vibration environment and to the dynamic stability. The stability suffers because these small tires normally have less cornering capability. The modular concept, now used in a growing number of countries within the European Union, offers a good solution to this problem [30,31]. The modular concept makes it possible to couple vehicle units based on modules defined in the EC directive 96/53 to longer combinations than normally allowed. This results in vehicle combinations with superior dynamic stability than current European vehicle combinations. When these new modular combinations shall be defined in regulations, it would be a good approach to use performance-based standards (PBS) as a basis [32]. The regulations have to be prescriptive in their nature but based on PBS. Figure 33.28 shows the result of the conversion of three current European vehicle combinations, two tractor–semitrailer combinations and one truck/center-axle combination, into two longer modular vehicle combinations.

Active vehicle stability control systems reduce the risk of instability significantly. This may be achieved by differential braking of the wheels, which is possible to do with electronic brake systems (EBS). Such systems assist the driver to avoid rollovers and yaw instability.

The rollover control intervenes when the lateral acceleration is approaching the static rollover threshold by cutting off the engine torque and using the wheel brakes to reduce the speed (see Figure 33.29). If a wheel lift is about to occur, full brake power is used.



FIGURE 33.28 Three vehicle combinations converted to two.

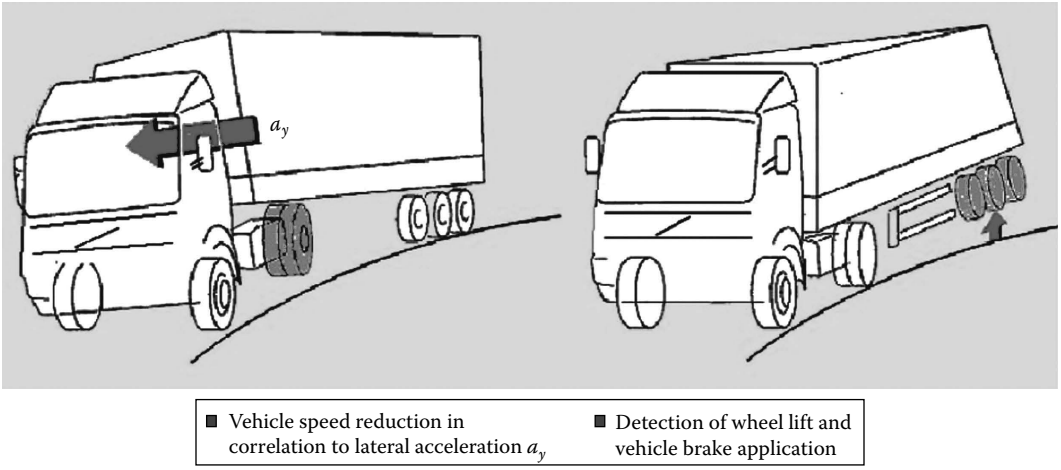


FIGURE 33.29 Rollover prevention.

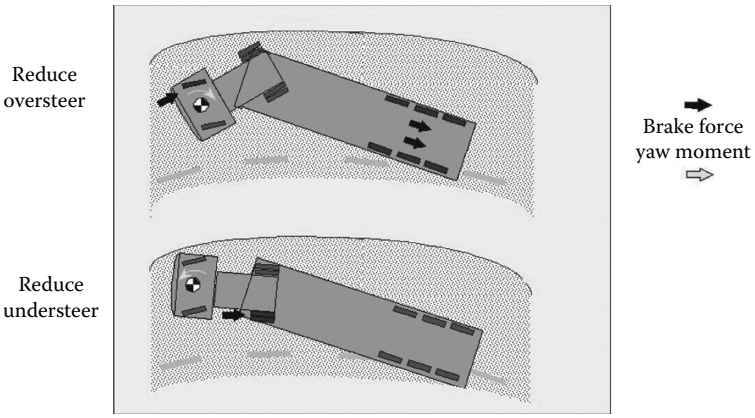


FIGURE 33.30 Yaw control.

The yaw control intervenes when the measured yaw behavior of the truck does not agree with the estimated yaw behavior that the truck should have for a certain combination of steering-wheel angle and speed. There may be two situations. One is that the yaw velocity of the tractor is too large, which leads to instability. The other is that the tractor is not turning enough, which may lead to that the vehicle cannot be kept on the road. In the first case, the outer front wheel of the tractor and the trailer wheels are braked in order to stop the yaw instability. In the second case, the rear inner wheel of the tractor is braked in order to increase the turning (see Figure 33.30).

33.10 PERCEPTION

The driver's perception of the truck is what matters at the end. Various control tasks such as path following, obstacle avoidance, or headway control have to be performed. The driver shall feel comfortable and relaxed also during long driving periods. It is important to understand that driver and vehicle together constitute a system. The driver acts as a controller getting feedback from a large number of sources and giving instructions via the controls to the vehicle. It is then important to understand on what stimuli the driver reacts. In order to make a mathematical model of the driver to include in a vehicle model, one must also understand how the driver reacts. There are a number of perception channels [33]. The most important one is vision. If the driver cannot see the road, for example, due to fog, they cannot drive. Another channel is motions of the vehicle, translational and rotational. They can be obtained from the inner ear or be body-distributed. The tactile input finally is from the controls, steering wheel, and pedals. The most important tactile channel is probably force and displacement of the steering wheel. It is important to know the correlation between the driver's perception of the vehicle behaviour and various vehicle dynamics characteristics. Driving simulators with large moving base are excellent tools for investigation of these relationships [34]. Figure 33.31 shows the high-performance driving simulators at VTI, Swedish National Road and Transport Research Institute.

Objective characteristic values obtained from directional response investigations correlate very well with the subjective perception of maneuver stability in lane changes and avoidance maneuvers. They are also a part of parameters that are relevant for the perception of on-center handling. Both the understeer gradient and the gain of yaw velocity with respect to steering-wheel angle have optimum values. The response time shall be as short as possible. The tactile sensory channels are very important for the perception of on-center handling. Among the more important relationships is the



FIGURE 33.31 VTI driving simulator.

relationship between steering-wheel torque and lateral acceleration. Both gradients and hysteresis correlate well with the subjective perception. Uncontrolled motions of the truck, primarily roll, are detrimental to the perception of on-center handling.

Steer by wire is currently used to steer rear axles. In the future, we may also see the front wheels steered by wire. This gives extended possibilities to tune the characteristics of the steering and optimize the tactile interface. It also gives the possibility to adapt the steering to the preferences of individual drivers. It is important to know how drivers are reacting on various stimuli, and it is also essential to have that knowledge in the context of active control systems and various driver-assist systems and how the driver's behavior is affected by these systems.

A driver is very adaptive to nonlinear dynamics, that is, they is a very robust controller. One example is that a driver, at least some drivers, can cope with an oversteered vehicle running above its critical speed, although it is unstable. Another example is the case with a blowout of a front tire. Although the dynamics of the truck changes significantly when one front tire loses almost all cornering stiffness, the drivers have no difficulty to adapt to the new characteristics and follow the prescribed path in all kinds of maneuvers, without and with braking. It is however also so that if the driver does not have full attention to the driving, there is a long reaction time before they takes action. Driving simulator tests have confirmed this. It was also shown that the reaction time is long only the first time a blowout happens. If it happens again during the same test run, the driver is prepared and acts accordingly [35].

33.11 VEHICLE MODELS

Mathematical modeling is fundamental for the understanding and the analysis of vehicle dynamics, both for vibration analysis and for the analysis of handling. It is important that the model covers what is relevant for the phenomenon that is to be analyzed. It is also an advantage if the model is simple and not more complex than necessary. It is, for example, sufficient if the model contains only the modes that are of interest. An unnecessary high degree of complexity makes it more difficult to make a correct model and to interpret the results. With current multi body system (MBS) software, it is easy to make large models with a lot of degrees of freedom (DOFs) without having to set up the equations, which is done by the software. An advantage is obviously that the equations have no errors, but the disadvantage is that the understanding of what happens may not be so good and the interpretation of the results is then more difficult [36,37].

Linear models have big advantages and are advantageous to use for qualitative studies because the analyses may be easily done in the frequency domain. Quite often, suspensions have highly nonlinear elements. Leaf springs, for example, have often a high amount of friction and dampers have nonlinear characteristics. Linear characteristics of these elements may however be obtained from a nonlinear analysis with nonstationary excitation. A linear model may then use these linearized characteristics for the linear analyses. In similar ways, the nonlinear characteristics of tires can be linearized and used for the analysis of the lateral dynamics [38].

Validation of models is always important, and in particular with large models, it is essential. There are various ways to do the validation. One is to use field measurements. This is the only alternative, and a good one, for handling models. For models with road input, the advantage is that the excitation is realistic, but the disadvantage is that it is difficult to obtain exact knowledge of the input. False conclusions may therefore be drawn from road measurements, for example, due to periodicities of the road excitation that give peaks in a spectrum, which may be misinterpreted as resonances. The so-called wheelbase filter is another phenomenon that filters the spectra and therefore causes new peaks or attenuation of resonance peaks [39]. Many mistakes have been done in the interpretation of measurement results due to lacking knowledge of the excitation mechanisms. The wheelbase-filter effect in a simple 2-DOF vertical-pitch model is exemplified in Figure 33.32. Correlated inputs mean that the inputs on front and rear axles have a constant time delay. This causes the wheelbase-filter effect if the speed is constant.

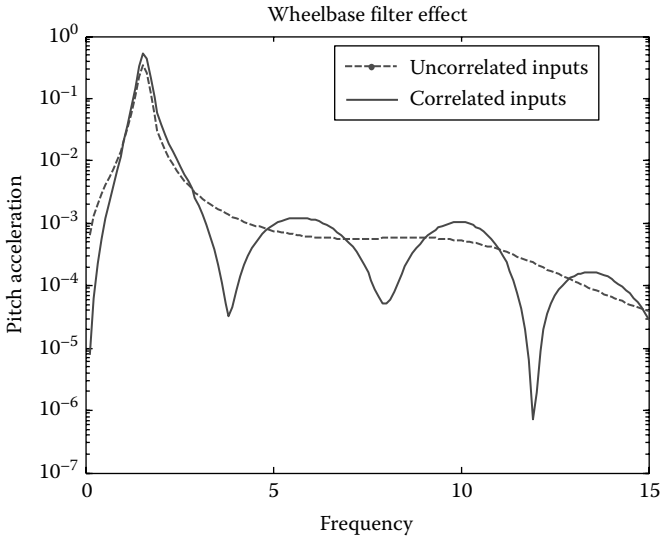


FIGURE 33.32 Wheelbase-filter effect.



FIGURE 33.33 Volvo road simulator.

An excellent way to validate a vehicle model with vertical input is to use a road simulator with vertical input on each wheel. Figure 33.33 shows the road simulator at Volvo in Gothenburg. It can also be used to visualize various phenomena. The road simulator measurements give very reproducible results and the excitation is well controlled and well known. In order to be able to correctly compare different vehicle configurations, independent of speed, the excitation on different axles shall be uncorrelated.

If the fatigue life of a front leaf spring is to be estimated, this can of course be done in a very elaborated vehicle model with a large number of DOFs. By studying what is influencing the spring force and the stress in the spring, it is possible to reduce the complexity. From linear models, it was seen that the influence from structural modes defined by frame flexibility, chassis packaging, and cab/engine installation was negligible. The wheelbase-filter effect can also be seen on the RMS

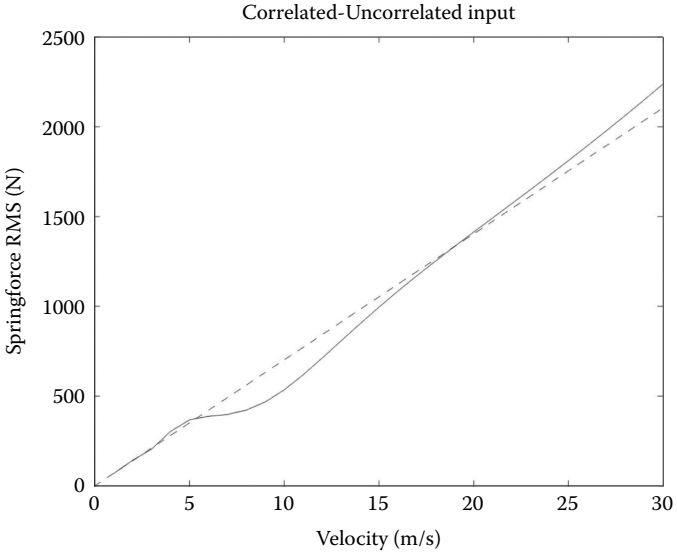


FIGURE 33.34 Wheelbase-filter effect on RMS values.

values of the spring force. Figure 33.34 shows that the RMS force is not proportional to the velocity (road severity), as it should be expected to be in this case, where the road has a waviness of 3.

The contributions from other axles than the front axle are however normally small. If the relationship between the squared radius of inertia and the product of the distances, front and rear, to the center of gravity, that is, the dynamic index, equals one, there is no contribution at all, because all the axle inputs are dynamically uncoupled. It is therefore sufficient to use a model with only front axle input. Figure 33.35 shows a comparison between the spring-force power spectral density (PSD) for a 2-DOF "quarter car" model (full line) and for a 7-DOF "half car" model (dashed line) with body and axle roll DOFs. The axle resonances differ (range 8–12 Hz), but as can be seen, the dynamic spring forces are mainly caused by low-frequency deflections, and the 2-DOF model may be used for a principal study. The nonlinear characteristics of the suspension are significant and are

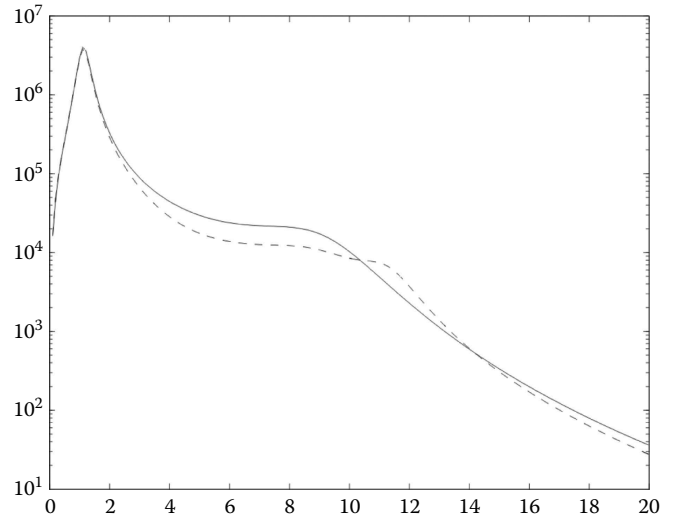


FIGURE 33.35 PSD from "quarter car" and "half car."

included, that is, a hysteresis of the leaf spring, a strongly nonlinear auxiliary spring as a bump stop, and a damper with nonlinear characteristic and elastic attachments.

The fatigue life of the front leaf spring is predicted by using a rainflow-counting algorithm on the simulated time responses. The reference spring is designed to have a maximum stress of twice the fatigue limit of the material. For comparison, a soft suspension in severe operations is used as reference.

The level-crossing distribution of the stress in Figure 33.36 shows the influence of the road description. Both the stationary and the nonstationary roads have the same RMS value and waviness. A nonstationary road consists of a mixture of good and bad roads, while a stationary road has constant road quality. It is not possible to obtain an adequate estimate of the fatigue life in the stationary case. The maximum stresses are less than the fatigue limit, which results in an infinite fatigue life [40].

Table 33.1 shows how the fatigue life is influenced by road description, type of model, operation severity, and spring stiffness.

The necessity of a nonlinear model is not as crucial as the nonstationary road description. The operation severity is obviously very important for the fatigue life. A very interesting fact is that a soft suspension not only improves the ride but also increases the fatigue life of the leaf spring.

For handling analysis, quite a few basic characteristics may be satisfactorily analyzed with a simple two-DOF model, linear or nonlinear. Also for qualitative analysis of vehicle combinations,

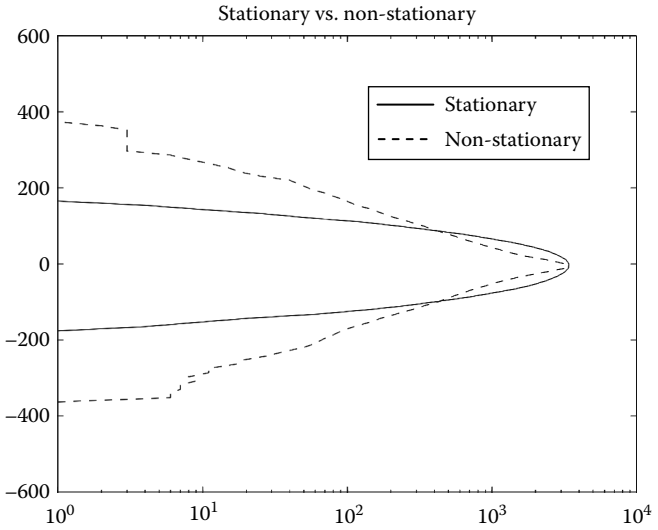


FIGURE 33.36 Level-crossing distributions of stress.

TABLE 33.1
Influence on Fatigue Life

Excitation		Model		Suspension		Operation		Relative
Stationary	Nonstationary	Linear	Nonlinear	Soft	Stiff +50%	Normal	Severe	Fatigue Life
x		x		x			x	∞
	x	x		x			x	1.45
	x		x	x			x	**1.00**
	x		x	x		x		19.8
	x		x		x		x	0.25

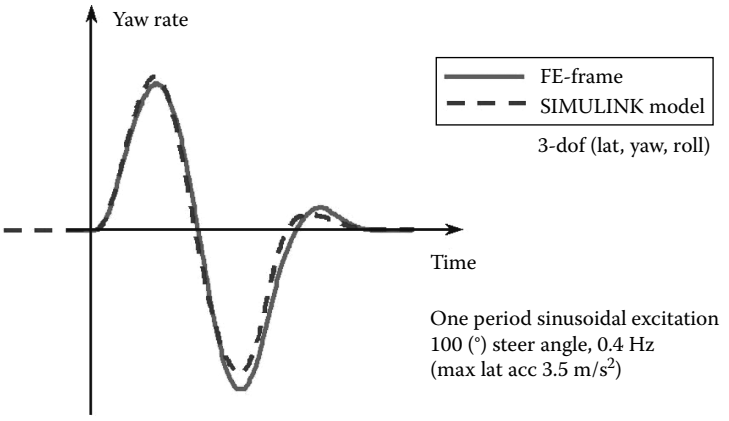


FIGURE 33.37 Responses from single sine-wave steer input.

it may be sufficient to use models, where each vehicle unit has just yaw and sideslip DOFs. The most important nonlinearity is the nonlinearity of the tires.

The example in Figure 33.37 shows the yaw velocity response from a single sine-wave steer input obtained from two different models of a rigid three-axle truck, one complex MBS model with flexible FE frame and one simple 3-DOF model. The comparison shows good agreement despite the large difference in complexity.

This illustrates that simple models may give sufficiently accurate results depending on what kind of phenomenon is to be analyzed. However, even when simple models are employed, they must reflect the real behavior of the truck.

One example is the flexible frame of the truck. It plays in many cases an important role in the dynamics of the truck, because it affects the load transfer left to right on the different axles. In a tractor, for example, the compliance of the frame must always be taken into account because no superstructure stiffens it. The frame should then not be modeled as a torsion model, with a torsion spring between two masses, as is often done, but as the warp model in Figure 33.38, which behaves in a similar way as the real frame [40].

It is difficult to understand how cab roll of a tractor is affected by stabilizers at different locations if a warp model is not used. Another example is a tipper with three-point attachment. This gives the effect that the resulting force that rolls the vehicle via the frame is moved rearward. The handling diagram in Figure 33.39 illustrates that the truck becomes oversteered at a rather low lateral acceleration, when this point moves away from the center of gravity. This behavior is due to the behavior of the frame and the effect of nonlinear tires.

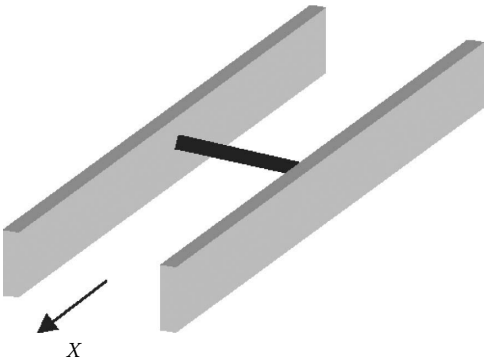


FIGURE 33.38 Warp model.

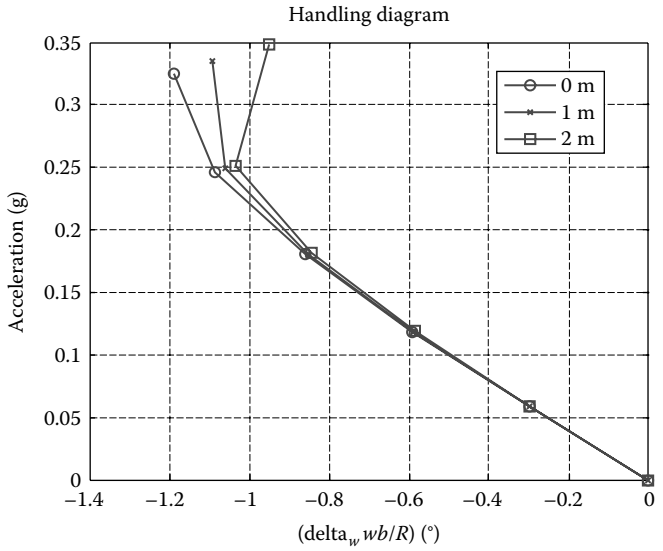


FIGURE 33.39 Steady-state responses with various superstructure attachments.

33.12 VERIFICATION AND VALIDATION

Verification of various vehicle dynamics feature characteristics is an important part in the product development process. A prerequisite for this is that there are feature requirements. The requirements consist of a set of quantified parameters for each feature area. It is obviously vital that these parameters are relevant for the perception of the feature they are measuring. This is a part of the validation. The verification is made with real prototypes but also with thoroughly verified vehicle models. One reason for using models in the verification process is to be able to verify a large number of different vehicle specifications used in different conditions.

For the verification of vehicle dynamics features, it is absolutely necessary to know in what environment and how the vehicle will be used. Figure 33.40 illustrates how the vehicle specification is affected by the operating environment, the vehicle utilization, and the transport mission. The optimal vehicle specification may be seen as an optimization of feature areas, weighted according to the preference of the customer.

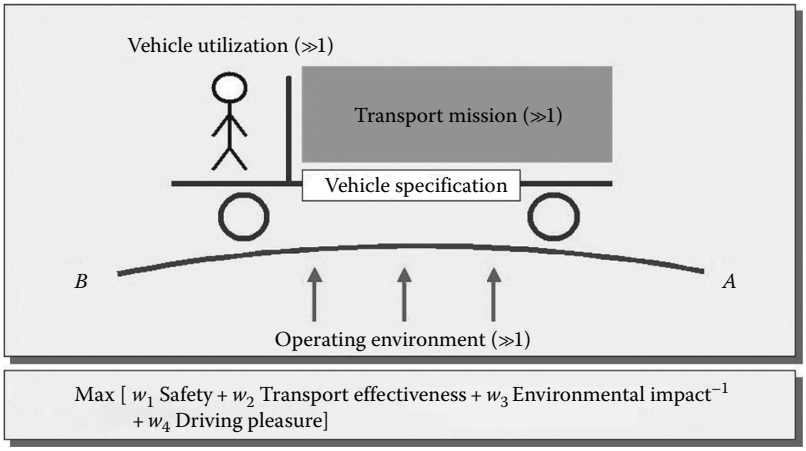


FIGURE 33.40 Factors affecting the vehicle specification.

TABLE 33.2
Examples of GTA-Parameters

Transport Cycle (Long-Distance Type, Distribution Type, Construction Type)			
Operative Environment		**Vehicle Utilization**	
Road condition	(Smooth, …., rough, …)	Speed changes	(Low, …, …, …)
Topography	(…, hilly, …)	Maneuvering	(Low, …, …, …)
Curve density	(Low, …, …)	Yearly usage	(…) kilometer/year or hours/year
Altitude	(…, 2000 m, …)	Transport mission	
Climate	(…, cold, …)	Body and equipment	(…, tank, tipper, …)
		GCW	(…, GCW20t, …, GCW200t)

A number of vehicle-independent parameters have been defined that specify differences in transport mission, operating environment, and vehicle utilization for haulage operations worldwide. These parameters define global truck application (GTA). Table 33.2 shows examples of GTA parameters.

The GTA parameters are high-level descriptions, which are a base that gives a uniform perspective on vehicle usage for all functions within the company such as product planning, product development, sales, and aftermarket. For the product development and for the feature verification in particular, the different parameter levels need to be quantified. The road condition is, for example, expressed in terms of Q-distribution and statistical distributions of the pothole content [41,42].

The quality of the verification results is an issue. There are a number of uncertainties related both to the description of the operation environment and to the quality of measurement and analyses (see Figure 33.41).

Also the features of a certain vehicle specification may vary from vehicle to vehicle depending on many factors, such as characteristic variations of vital components and the state of wear. All these result in a certain dispersion of the vehicle performance in various feature areas. This leads to that the targets of the vehicle performance have to be higher than the customer expectation in order to avoid problems. This may be called a probabilistic design (see Figure 33.42).

It is obviously necessary that the targets are objectively quantifiable parameters. Subjective assessments of vehicle dynamics features are however still rather frequent. There is a larger acceptance of objective ride measurements than of objective handling measurements. Much of the vehicle dynamics research has been focused on the driver's perception of various aspects of vehicle dynamics, and there is enough knowledge today to be able to rely entirely on objective measurements. Subjective assessments are often misleading and therefore detrimental to the development process.

The repeatability also among professional truck drivers is poor. Figure 33.43 shows results from a well-controlled driving simulator test in a histogram. The same vehicle was assessed twice and the

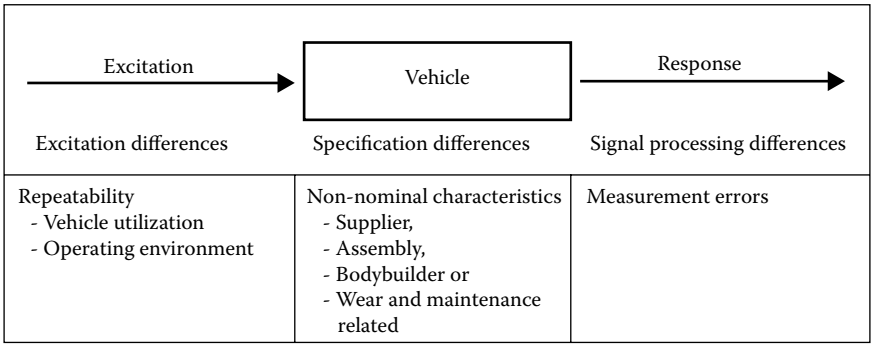


FIGURE 33.41 Quality of the verification process.

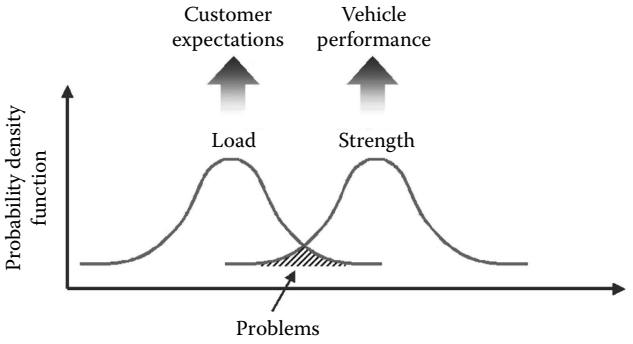


FIGURE 33.42 Matching expectations and performance.

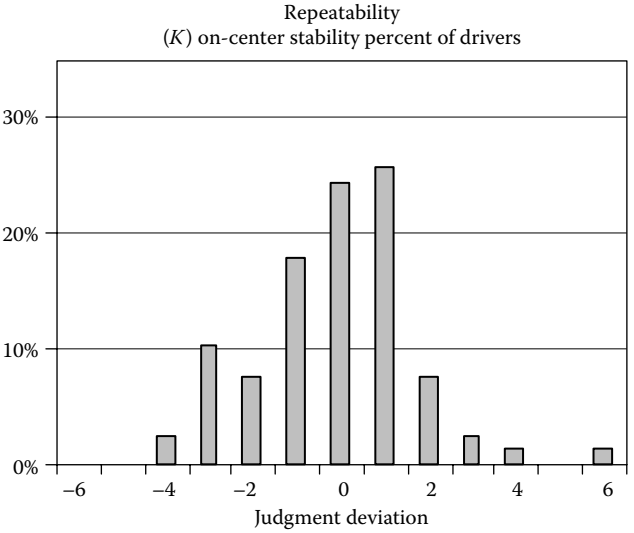


FIGURE 33.43 Result from repeatability test.

histogram shows how the assessment deviations are distributed. Only 25% of the drivers managed to give the same judgment each time. Sixty-eight percent stayed within 1 point's deviation and quite a few were completely unable to recognize the identical vehicle. The assessment scale had seven levels.

In another test, carried out on public roads, the drivers (A–K) were asked to judge the overall handling behavior at two occasions. Four of the drivers made three test drives. There were two test

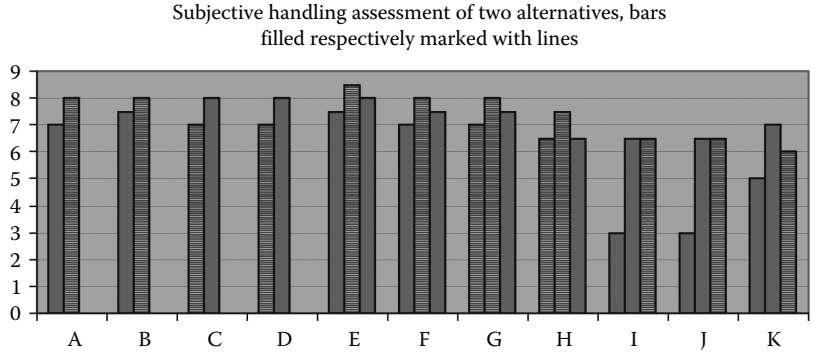


FIGURE 33.44 Results from handling assessment.

alternatives, and between the test runs, the vehicle was modified or said to be modified. The two test alternatives were driven in different order by the different drivers. The result from this test is shown in Figure 33.44. It appears that the order between test runs and assessment is the only correlation there is. The second test run always gives the highest rating regardless of which alternative it is. This illustrates the unreliability of subjective assessment.

33.13 CONCLUSION

There has been a tremendous development of heavy vehicles and of the dynamics of heavy vehicles during the latter half of the last century. The drivers for this development have been the increasing demands on transport effectiveness, the need of specialized and optimized vehicles, the requirements on better active safety, and better vibration environment.

The most important in the development process is to understand all aspects of the dynamics and to set up objective requirements.

Mathematical modeling is fundamental, but unnecessary complexity should be avoided.

Verification should be carried out both in real vehicles and with mathematical models and shall be based on objective criteria and not on subjective assessment.

Active control systems to assist the driver are developing quickly and increase the safety provided that they are added to a truck with good passive features.

REFERENCES

1. EU DIRECTIVE 96/53/EC laying down for certain road vehicles circulating within the Community the maximum authorized dimensions in national and international traffic and the maximum authorized weights in international traffic, 25 July 1996.
2. A. Grimm, C.B. Winkler, and R. Sweet, *Bibliography of Heavy Vehicle Dynamics*, University of Michigan Transportation Research Institute, Ann Arbor, MI, June 1999.
3. L. Segel, P. Fancher, C. Winkler, R. Ervin, T. Gillespie, and M. Sayers, Mechanics of heavy-duty truck systems, Course notes. Vol. 1 and 2, University of Michigan Transportation Research Institute, Center for Professional Development, Ann Arbor, MI. 2002.
4. J. Aurell and S. Edlund, Vehicle dynamics of commercial vehicles, Volvo Technology Report, 1990.
5. J. Aurell and T. Wadman, Impact of European size and weight policies on the characteristics of heavy vehicles, *Proceedings of the Fourth International Symposium on Heavy Vehicle Weights and Dimensions*, Ann Arbor, MI, June 25–29, 1995.
6. J. Aurell, Effects of infrastructure on vehicles, *DIVINE Concluding Conference*, Ottawa, Ontario, Canada, June 23–25, 1997.
7. J. Aurell and S. Edlund, Identifiering och Simulering av Vägstörningar mha Vägsimulator, *Proceedings from SVIB's Special Symposium om Vibrationer och ljud I farkoster*, Riksgränsen, Sweden, 1990.
8. J. Aurell and S. Edlund, Olika Typer av Vägexcitering på Lastbilar, *Proceedings from SVIB's special symposium om Nyare metoder för mätning och analys av vibrationer i maskiner och fordon*, 1987.
9. J. Ribarits, J. Aurell, and E. Andersers, Ride comfort aspects of heavy truck design, SAE 781067, 1978.
10. ISO 2631–1:1997, Mechanical vibration and shock-Evaluation of human exposure to whole-body vibration-Part 1: General requirements.
11. EU DIRECTIVE 2002/44/EC on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (Vibration), June 25 2002.
12. J. Aurell, On the influence of different suspension parameters on dynamic wheel loads, *Proceedings from ImechE seminar on Road Wear: The Interaction between Vehicle Suspensions and the Road*, London, U.K., 1991.
13. DIVINE, Dynamic interaction between vehicles and infrastructure experiment Technical report. OECD Report DST/DOT/RTR/IR6(98)1/FINAL.
14. COST 334, Effects of wide single tyres and dual tyres, European Commission Directorate General Transport, 2001.

15. J. Aurell, Some views of the truck industry on wide single tyres on driving axles, *Proceedings of the Seventh International Symposium on Heavy Vehicle Weights and Dimensions*, Delft, the Netherlands, June 16–20, 2002.
16. P.S. Fancher, R.D. Ervin, C.C. MacAdam, and C.B. Winkler, Measurements and representation of the mechanical properties of truck leaf springs, SAE 800905, 1980.
17. ISO 8608:1995, Mechanical vibration—Road surface profiles—Reporting of measured data.
18. J. Aurell and S. Edlund, Operating severity distribution—A base for vehicle optimisation, *Proceedings of eleventh IAVSD-Symposium in Kingston*, Kingston, Ontario, Canada, August 21–25, 1989.
19. F. Öijer, FE-based vehicle analysis of heavy trucks part II: Prediction of force histories for fatigue life analysis, *Proceedings of the second MSC Worldwide Automotive Conference*, Dearborn, MI, 2000.
20. C.B. Winkler and J. Aurell, Analysis and testing of the steady-state turning of multiaxle trucks, *Proceedings of the Fifth International Symposium on Heavy Vehicle Weights and Dimensions*, Queensland, Australia, March 29–April 2, 1998.
21. ISO 14793, Road vehicle—Heavy commercial vehicles and buses—Lateral transient response test methods.
22. M. Mitschke and H. Wallentowitz, *Dynamik der Kraftfahrzeuge*, Springer, Berlin, Germany, 2006.
23. VERTEC, Vehicle, road, tyre and electronic control system interaction—Prediction and validation of handling behaviour, EC-Contract G3RD-2002-0805, *Third International Colloqium on Vehicle Tyre Road Interaction*, Stuttgart, Baden-Württemberg, Germany, March 8, 2006.
24. J. Aurell and S. Edlund, The influence of steered axles on the dynamic stability of heavy vehicles, SAE 892498, 1989.
25. B.A. Jujnovich, *Active Steering of Articulated Vehicles*, Cambridge University Engineering Department, Cambridge, U.K., 2005.
26. C.B. Winkler, D. Blower, R.D. Ervin, and R.M. Chalasani, Rollover of heavy commercial vehicles, SAE Research Report, 2000.
27. J. Aurell and C.B. Winkler, Standard test procedures for the lateral stability of heavy vehicle combinations, *Proceedings of the Fourth International Symposium on Heavy Vehicle Weights and Dimensions*, Ann Arbor, MI, June 25–29, 1995.
28. J. Aurell and J. Koppenaal, Simplified procedure for determining lateral stability of heavy vehicle combinations, *Proceedings of the Fifth International Symposium on Heavy Vehicle Weights and Dimensions*, Queensland, Australia, March 29–April 2, 1998.
29. ISO 14791, Road vehicles—Heavy commercial vehicle combinations and articulated buses—Lateral stability test methods.
30. J. Aurell and T. Wadman, Vehicle combinations based on the modular concept, Report no. 1/2007, Committee 54, Vehicles and Transports, Nordic Road Association.
31. L. Pilskog, J. Aurell, and C. Avedal, Experience from the European modular system in Scandinavia, *DEKRA/VDI Symposium Safety of Commercial Vehicles*, Neumünster, Schleswig-Holstein, Germany, October 12–13, 2006.
32. J. Aurell, Potential implications of PBS for the transport industry, *Proceedings from PBS International Seminar*, Melbourne, Australia, February 10–12, 2003.
33. C.C. MacAdam, Understanding and modeling the human driver, *Vehicle System Dynamics*, 40(1–3), 101–134, 2003.
34. J. Aurell and N. Fröjd, Correlation between objective handling characteristics and subjective perception of handling qualities of heavy vehicles, *Proceedings from Fifth International Symposium on Advanced Vehicle Control*, Ann Arbor, MI, August 22–24, 2000.
35. H. Pettersson, J. Aurell, and S. Nordmark, Truck driver behaviour in critical situations and the impact of surprise: A pilot study of a sudden blow-out on the front axle of a heavy truck, DSC 2006, Paris, France, October 2006.
36. I. Johansson and S. Edlund, Optimization of vehicle dynamics in trucks by use of full vehicle models, *Proceedings from ImechE Conference on Vehicle Ride and Handling*, Birmingham, U.K., 1993.
37. F. Öijer and S. Edlund, Identification of transient road obstacle distributions and their impact on vehicle durability and driver comfort, *Eighteenth IAVSD Symposium*, Kanagawa, Japan, August 24–30, 2003.
38. J. Aurell, Models for assessment of suspensions with regard to dynamic wheel forces, *Engineering Foundation symposium on Vehicle-Road and Vehicle-Bridge Interaction*, Noordwijkerhout, the Netherlands, 1994.
39. J. Aurell and S. Edlund, Prediction of vibration environment in real operations described by Q-distributions, *Proceedings of 14th IAVSD-Symposium*, Ann Arbor, MI, August 21–25, 1995.

40. J. Aurell, The influence of warp compliance on the handling and stability of heavy commercial vehicles, *Proceedings from Sixth International Symposium on Advanced Vehicle Control*, Hiroshima, Japan, September 9–13, 2002.
41. S. Edlund and P.-O. Fryk, The right truck for the job with global truck application descriptions, SAE 2004-01-2645, 2004.
42. S. Edlund and F. Öijer, Complete vehicle durability assessments using discrete sets of random roads and transient obstacles based on Q-distributions, Seventeenth IAVSD-Symposium, Lyngby, Denmark, 2001.
43. ISO 14792, Road vehicles—Heavy commercial vehicles and buses—Steady-state circular tests.

34 Dynamics of Off-Road Vehicles

Jo Y. Wong

CONTENTS

34.1 Introduction 1279
34.2 Performance (Mobility) of Off-Road Vehicles 1280
34.2.1 Computer Simulation Model NTVPM for Vehicles with Flexible Tracks 1281
34.2.1.1 Prediction of Normal Pressure Distribution under the Track 1281
34.2.1.2 Prediction of Shear Stress Distribution under the Track 1283
34.2.1.3 Prediction of Motion Resistance, Thrust, and Drawbar Pull 1284
34.2.1.4 Experimental Substantiation 1285
34.2.1.5 Applications 1286
34.2.2 Computer Simulation Model RTVPM for Vehicles with Rigid Link Tracks 1286
34.2.2.1 Basic Approach to the Development of the Model 1287
34.2.2.2 Experimental Substantiation 1288
34.2.2.3 Applications 1288
34.2.3 Computer Simulation Model NWVPM for Off-Road Wheeled Vehicles 1289
34.2.3.1 Basic Approach to the Development of the Model 1290
34.2.3.2 Experimental Substantiation 1290
34.2.3.3 Applications 1291
34.3 Handling Characteristics of Off-Road Vehicles 1291
34.3.1 Steering of Tracked Vehicles 1292
34.3.2 Simplified Analysis of the Mechanics of Skid-Steering for Tracked Vehicles 1293
34.3.3 General Theory for Skid-Steering 1295
34.3.3.1 Basic Approach to the Development of the General Theory 1296
34.3.3.2 Experimental Substantiation 1298
34.3.3.3 Coefficient of Lateral Resistance 1300
34.3.4 Steering of Off-Road Wheeled Vehicles 1302
34.4 Ride Characteristics of Off-Road Vehicles 1304
34.4.1 Ride Comfort Criteria 1304
34.4.2 Characterization of Surface Irregularities 1305
34.4.3 Modeling Ride Dynamics of Off-Road Vehicles 1306
34.5 Closing Remarks 1309
Acknowledgments 1310
References 1310

34.1 INTRODUCTION

The dynamic behavior of an off-road vehicle may be described in terms of its performance (mobility), handling, and ride. Performance characteristics refer to the ability of the vehicle to overcome motion resistance, to develop drawbar pull, to negotiate slope, to accelerate, and to decelerate. Handling behavior is primarily concerned with the steering response of the vehicle to the driver's commands. Ride quality is related to the vibration of the vehicle excited by surface irregularities and its effects on the driver and passenger.

There are currently two basic forms of running gear for off-road vehicles: wheel and track. While both off-road wheeled vehicles and road vehicles use tires as their running gear, their functional requirements are significantly different. Off-road wheeled vehicles are designed for operation over unprepared terrain, which may have significant deformation under vehicular loads, whereas road vehicles are designed for operation over paved surfaces, which have relatively little deformation in comparison with unprepared terrain under similar loading conditions. For heavy off-road vehicles, tracks are usually employed as the running gear, primarily to obtain larger contact area in order to reduce the contact pressure on the terrain. It is interesting to note that the concept of the track was first conceived in the late eighteenth century as a "portable railway" to wheel carriages. This "portable railway" is laid down on the terrain in front of the road wheels and picked up at the rear as the vehicle advances. As the track system has certain unique features and tracked vehicles are widely used in off-road operations, in this chapter, greater emphasis will be placed on the discussion of the dynamics of tracked vehicles.

34.2 PERFORMANCE (MOBILITY) OF OFF-ROAD VEHICLES

Dependent upon the functional requirements, different criteria are employed to evaluate the performance characteristics of various types of off-road vehicles. For tractors, their main function is to provide draft to pull various types of implement or machinery. Consequently, their drawbar performance is of prime interest. It may be characterized by the ratio of drawbar pull to vehicle weight (drawbar pull coefficient), drawbar power (the product of drawbar pull and vehicle forward speed), and drawbar efficiency (the ratio of drawbar power to the corresponding power delivered by the driven wheels or sprockets). For cross-country transport vehicles, the transport productivity (the product of payload and vehicle forward speed) and transport efficiency (the ratio of transport productivity to the corresponding power delivered by the power plant of the vehicle) are often used as the basic criteria for evaluating their performance. For military vehicles, on the other hand, the maximum feasible operating speed between two specific points in a given region may be employed as a criterion for evaluating their mobility.

Although differing criteria are used to assess the performance of different types of off-road vehicles, there is a basic requirement common to all cross-country vehicles, that is, their mobility over unprepared terrain. Mobility in a broad sense refers to the capability of the vehicle in traversing soft terrain, in negotiating obstacles, in providing comfortable ride over rough terrain, and in water crossing. The ability in traversing soft terrain is a basic concern in vehicle mobility, and detailed analysis of the relationship among vehicle performance, vehicle design parameters, and terrain characteristics is, therefore, of prime importance.

In this chapter, emphasis will be placed on the discussion of the latest developments in computer simulation models for performance evaluation of off-road vehicles, including both wheeled and tracked vehicles.

As noted in Chapter 20, in the past owing to the lack of understanding of the response of natural terrain to vehicular loading and of the complex phenomenon of vehicle–terrain interaction, empirical and semiempirical methods have been widely used in the evaluation or prediction of off-road vehicle performance (mobility). In recent years, with the increasing globalization of the economy, improving competitiveness in the world market has become a major concern of industry. The empirical approach to product development adopted in the past has become extremely inefficient and prohibitively expensive. To gain a competitive edge, shortening the product development cycle is of vital importance. Virtual product development (virtual prototyping) is therefore being adopted by industry. To implement this process successfully, the development of realistic and comprehensive computer simulation models for performance and design evaluation of off-road vehicles is essential. In response to this requirement, a series of computer simulation models for tracked and wheeled vehicles has been developed [1,2]. In contrast with empirical or semiempirical methods developed earlier, these computer simulation models are based on the detailed analysis of the physical nature

of vehicle–terrain interaction and on the principles of applied mechanics. They take into account all major design features of the vehicle and its running gear. All pertinent characteristics of the terrain are also taken into consideration, including the pressure-sinkage and shearing characteristics, as well as the response to repetitive loading, as described in Section 20.6.1.

For evaluating the performance of vehicles with rubber-belt tracks or link tracks with relatively short track pitch, commonly found in military fighting, and logistics vehicles, a model known as NTVPM has been developed [3–9]. In this model, the track is idealized as a flexible and extensible belt. For tracked vehicles with rigid link tracks having relatively long track pitch, commonly used in farming and construction industry, a model known as RTVPM has been developed [10–12]. In this model, the track is considered as a system of rigid links connected through pins. For off-road wheeled vehicles, including wheeled armored vehicles and tractors, a model known as NWVPM has been developed. This model takes into account all major design features of the vehicle, as well as the tire. Multipass effects of the succeeding tires running in the ruts formed by the preceding tires are taken into consideration.

34.2.1 Computer Simulation Model NTVPM for Vehicles with Flexible Tracks

For high-speed tracked vehicles, such as military fighting and logistics vehicles and cross-country transport vehicles, rubber-belt tracks or link tracks with relatively short track pitch are commonly used. This kind of short-pitch track system typically has a ratio of road wheel diameter to track pitch in the range of 4–6, a ratio of road wheel spacing to track pitch ranging from 4 to 7, and a ratio of sprocket pitch diameter to track pitch of the order of 4. The rubber-belt track and short-pitch link track will be referred to as "flexible tracks," and they are idealized as a flexible and extensible belt in the analysis of track–terrain interaction.

The computer simulation model NTVPM has been developed for performance and design evaluation of tracked vehicles with flexible tracks. It takes into account all major design parameters of the vehicle, including vehicle weight, track system configuration, number of road wheels, road wheel spacing, track dimensions and geometry, initial track tension, track longitudinal elasticity, suspension characteristics, location of the center of gravity, arrangements for the sprockets, idlers, and supporting rollers, and vehicle hull (belly) shape (for the analysis of vehicle hull–terrain interaction, when the track sinkage is greater than the ground clearance and the vehicle hull is in contact with the terrain surface). All pertinent terrain characteristics, including the pressure–sinkage and shearing characteristics and the response to repetitive loading, are taken into consideration [1–9].

NTVPM can be used to predict the normal and shear stress distributions on the track–terrain interface and the external motion resistance, thrust, drawbar pull, and tractive efficiency of the vehicle as functions of track slip under steady-state operating conditions. It is particularly suited for the evaluation of competing designs, for the optimization of vehicle design parameters, and for the selection of vehicle candidates from tractive performance standpoint. The basic features of NTVPM have been validated by means of full-scale vehicle tests on a variety of terrains, including mineral terrain, organic terrain, and snow-covered terrain.

The basic approach to the development of NTVPM is outlined in the succeeding text [1–3].

34.2.1.1 Prediction of Normal Pressure Distribution under the Track

As noted previously, in the development of NTVPM, the track is treated as a flexible and extensible belt. A schematic of the track–road wheel system on a deformable terrain under steady-state operating conditions is shown in Figure 34.1.

When a tracked vehicle rests on a hard surface, the track lies flat on the ground. In contrast, when the vehicle travels over a deformable terrain, the normal load applied through the track system causes the terrain to deform. The track segments between road wheels take up load, and as a result, they deflect and have the form of a curve. The actual length of the track in contact with the terrain between the front and rear road wheels increases in comparison with that when

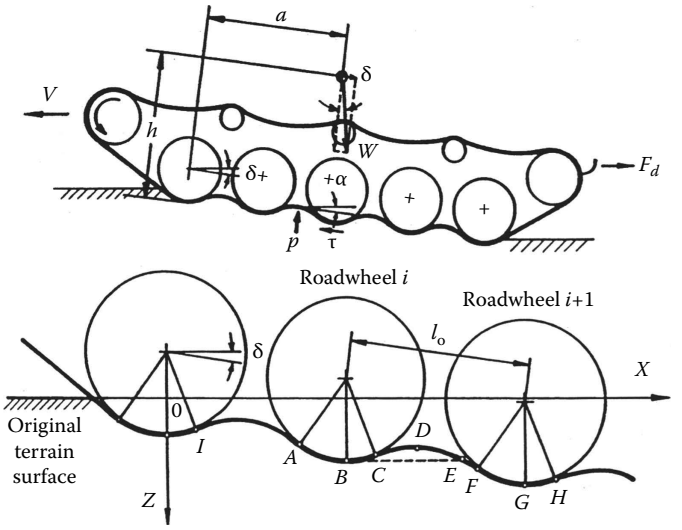


FIGURE 34.1 Interaction between a flexible track and a deformable terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

the track rests on a firm ground. This causes a reduction in the sag of the top run of the track and a change in track tension. In the model, the elongation of the track under tension is taken into consideration.

The deflected track in contact with the terrain may be divided into two sections: one in contact with both the road wheel and the terrain, such as *AC* and *FH* shown in Figure 34.1, and the other in contact with the terrain only, such as segment *CF*. The shape of the track segment in contact with the road wheel, such as *AC*, is defined by the shape of the road wheel, whereas the shape of the track segment in contact with the terrain only, such as *CF*, is determined by track tension, road wheel spacing, pressure–sinkage relationship, and response to repetitive loading of the terrain.

Along segment *AB*, the pressure exerted on the terrain increases from *A* to *B*. From *B* to *D*, the pressure decreases corresponding to the unloading portion of the repetitive loading cycle shown in Figure 20.15. Along segment *DE*, the pressure increases again, corresponding to the reloading portion of the repetitive loading cycle. Beyond point *E*, which is at the same level as point *B*, the sinkage is higher than that at *B*. As a result, the pressure increases and the sinkage of the succeeding road wheel will be greater than that of the preceding road wheel. This causes the vehicle to assume a nose-up attitude. Beyond point *G*, the pressure exerted on the terrain decreases again and another unloading–reloading cycle begins.

Over highly compressible terrain, such as fresh, deep snow, track sinkage may be greater than the ground clearance of the vehicle. If this occurs, the hull (belly) of the vehicle will be in contact with the terrain surface and will support part of the vehicle weight. This will reduce the load carried by the tracks and will adversely affect the traction of the vehicle on frictional terrain. Furthermore, the contact of the hull with the terrain will give rise to an additional drag component, the belly drag. The characteristics of vehicle hull–terrain interaction and their effects on vehicle performance have been taken into consideration in NTVPM. In the model, the characteristics of the independent suspension of the road wheels are fully taken into account. Torsion bar suspensions, hydropneumatic suspensions with nonlinear characteristics, and others can be accommodated in the model.

Based on the understanding of the physical nature of track–terrain interaction described earlier, a set of equations for the equilibrium of forces and moments acting on the track–road wheel system and for the evaluation of the overall track length is derived. They establish the relationship among the shape of the deflected track in contact with the terrain, vehicle design parameters, and terrain characteristics.

The solution to this set of equations defines the sinkage of the road wheels, inclination of the vehicle body, track tension, and track shape in contact with the terrain. From these and taking into account the pressure–sinkage relationship and response to repetitive loading of the terrain, the normal pressure distribution under a tracked vehicle in steady-state motion is predicted.

34.2.1.2 Prediction of Shear Stress Distribution under the Track

The tractive performance of a tracked vehicle is closely related to the normal pressure and shear stress distributions on the track–terrain interface. As discussed in Section 20.6.1.2, shear stress is related to shear displacement. To predict the shear stress distribution, the development of shear displacement under the track should be determined first. The shear displacement under a flexible track may be evaluated by analyzing the slip velocity V_j on the track–terrain interface. The slip velocity of a point on the flexible track relative to the terrain surface is the tangential component of the absolute velocity V_a shown in Figure 34.2. The magnitude of the slip velocity V_j is given by

$$V_j = V_t - V\cos\alpha = r\omega - r\omega(1-i)\cos\alpha = r\omega[1-(1-i)\cos\alpha] \tag{34.1}$$

where

r and ω are the pitch radius and angular speed of the sprocket
i is the slip of the track
α is the angle between the tangent to the track at point P and the horizontal
V_t is the theoretical speed of the vehicle (i.e., $V_t = r\omega$)
V is the actual forward speed of the vehicle

Based on the analysis of slip velocity, the shear displacement j along the track–terrain interface is given by

$$j = \int_o^t r\omega[1-(1-i)\cos]dt = \int_o^l r\omega[1-(1-i)\cos\alpha]\frac{dl}{\omega} = l-(1-i)x \tag{34.2}$$

where

l is the distance along the track between point P and the front contact point of the track
x is the corresponding horizontal distance between point P and the front contact point of the track

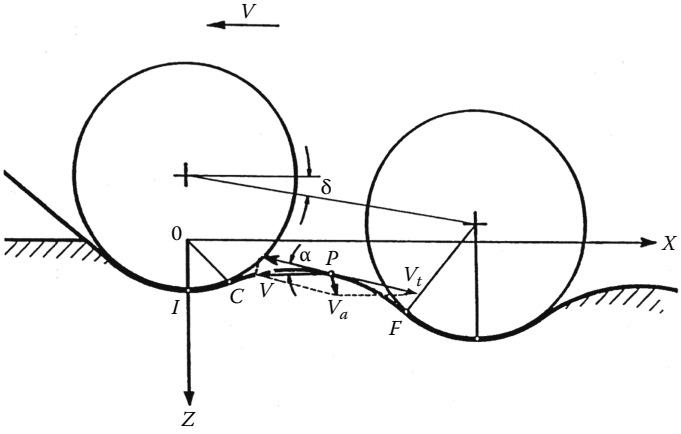


FIGURE 34.2 Slip velocity of a point on a flexible track in contact with a deformable terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

If the shear stress–shear displacement relationship is described by the exponential equation (Equation 20.22), the shear stress along the flexible track $\tau(x)$ is given by

$$\tau(x) = [c + p(x)\tan\phi]\left[1 - \exp\left(-\frac{l-(1-i)x}{K}\right)\right] \tag{34.3}$$

where $p(x)$ is normal pressure on the track, which is a function of the horizontal distance x.

It should be noted that shear stress is related to normal pressure. At the beginning of the simulation process, an initial estimate of the normal pressure distribution is made by neglecting the presence of shear stress on the track–terrain interface, using the procedure described in the preceding section. After the initial normal pressure distribution is determined, the procedure outlined earlier is used to estimate the shear stress distribution. An iterative process is then initiated to recalculate the normal pressure distribution based on the estimated shear stress distribution. The shear stress distribution is subsequently revaluated, based on the recalculated normal pressure distribution. This iterative procedure will be repeated until the errors in the equilibrium equations for forces and moments acting on the vehicle are less than the prescribed values.

34.2.1.3 Prediction of Motion Resistance, Thrust, and Drawbar Pull

When the normal and shear stress distributions under a tracked vehicle at a given slip have been determined, the tractive performance of the vehicle can be predicted. The performance of an off-road vehicle is usually characterized by its motion resistance, thrust, and drawbar pull (the difference between thrust and motion resistance) as functions of slip.

The external motion resistance R_c of the track can be determined by integrating the horizontal component of the normal pressure acting on the track–terrain interface. For a vehicle with two tracks, R_c is given by

$$R_c = 2b\int_o^l p\sin\alpha\, dl \tag{34.4}$$

where

b is the contact width of the track
l is the contact length of the track
p is the normal pressure
α is the angle between a track element on the interface and the horizontal

The thrust of the vehicle can be calculated by integrating the horizontal component of the shear stress on the track–terrain interface. For a vehicle with two tracks, the thrust F is given by

$$F = 2b\int_o^l \tau\cos\alpha\, dl \tag{34.5}$$

where τ is the shear stress on the track–terrain interface.

It should be noted that the thrust calculated by Equation 34.5 is due to the shearing action of the track across the grouser tips. For a track with high grousers, additional thrust will be developed from the shearing action on the vertical surfaces on either side of the track, and this should be taken into consideration [1,2].

The drawbar pull F_d is the difference between the thrust and the external motion resistance and is expressed by

$$F_d = F - R_c \tag{34.6}$$

It should be noted that if the track sinkage is greater than the ground clearance and the vehicle hull is in contact with the terrain surface, the hull (belly) drag should be taken into account in the calculation of the drawbar pull [1,2].

34.2.1.4 Experimental Substantiation

The basic features of NTVPM have been validated with field test data of both single-unit and two-unit articulated tracked vehicles over a variety of unprepared terrains, including mineral terrain, organic terrain, and snow-covered terrain. As an example, Figures 34.3 and 34.4 show a comparison of the measured and predicted normal pressure distribution under the track pad and that of the measured and predicted drawbar performance of an armored personnel carrier over an organic terrain, respectively. More validation test results may be found in references [1–3].

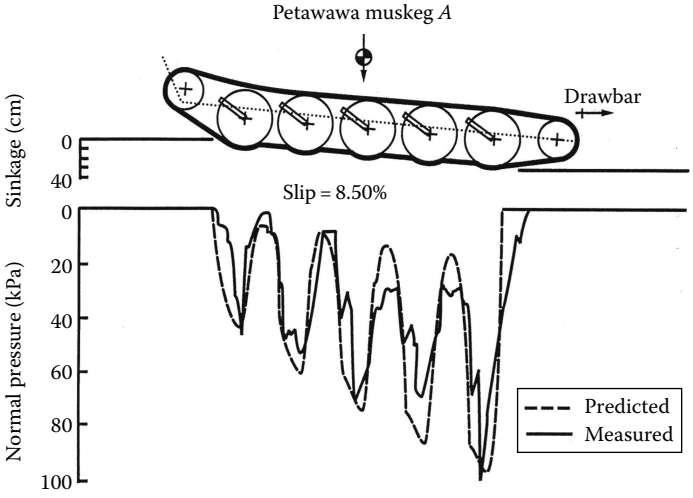


FIGURE 34.3 Comparison of the measured normal pressure distribution under the track pad of an armored personnel carrier and the predicted one using the computer simulation model NTVPM on an organic terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

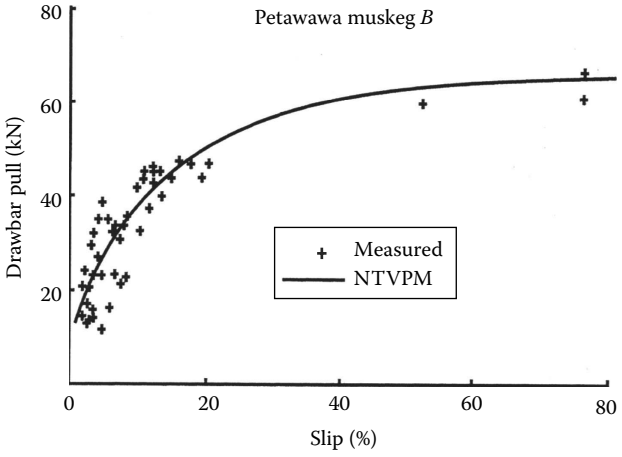


FIGURE 34.4 Comparison of the measured drawbar performance of an armored personnel carrier and the predicted one using the computer simulation model NTVPM on an organic terrain. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

34.2.1.5 Applications

NTVPM can be employed to evaluate the effects of vehicle design and terrain conditions on vehicle performance. It can also be used in parametric analysis and design optimization. As an example, Figure 34.5 shows the effects of the number of road wheels and initial track tension on the drawbar pull to weight ratio of an armored personnel carrier over deep snow predicted by NTVPM [1,2,4]. It can be seen that both the number of road wheels and initial track tension have significant effects on vehicle mobility over soft terrain. For a given vehicle, its performance over soft terrain can be greatly improved by appropriate increase of the initial track tension in most cases, dependent upon suspension settings [1,2,4,5]. This research finding from the applications of NTVPM has led to the development of a driver controlled, central initial track tension regulating system. Over normal terrain, the driver sets the initial track tension at the regular level. However, when traversing soft terrain is anticipated, the driver can readily increase the initial track tension to an appropriate level to improve vehicle mobility. This system has been introduced into a number of high-mobility military fighting vehicles. The effects of the central track tension regulating system on tracked vehicle mobility are analogous to those of the central tire inflation system on wheeled vehicle mobility.

NTVPM has been successfully used to assist vehicle manufacturers in the development of new products and governmental agencies in the selection of vehicle candidates in North America, Europe, Asia, and Africa.

34.2.2 Computer Simulation Model RTVPM for Vehicles with Rigid Link Tracks

For low-speed tracked vehicles, such as those used in farming and construction industry, rigid link tracks with relatively long track pitch are commonly used. This type of track system has a ratio of road wheel diameter to track pitch as low as 1.2 and a ratio of road wheel spacing to track pitch typically 1.5. This type of track will be referred to as "rigid link track" and may be treated as a system of rigid links connected with pins.

The computer simulation model RTVPM has been developed for performance and design evaluation of tracked vehicles with rigid link tracks. It takes into account all major design parameters of the vehicle, including vehicle weight, track system configuration, number of road wheels, road wheel spacing, track dimensions and geometry, initial track tension, location of the center of gravity, arrangements for the sprockets, idlers, and supporting rollers, and drawbar hitch location. As the

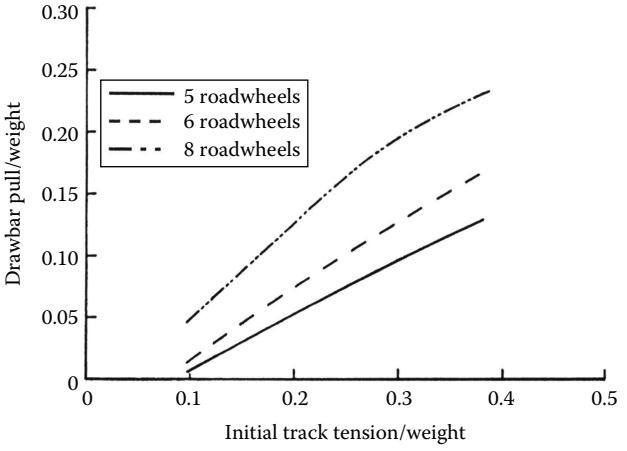


FIGURE 34.5 Effects of the number of road wheels and initial track tension to weight ratio on the drawbar pull to weight ratio of a tracked vehicle on snow predicted using the computer simulation model NTVPM. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 3rd edn., John Wiley, New York, 2001.)

track links are considered to be rigid, the track is regarded as inextensible [1,2,10–12]. Terrain parameters used in this model are the same as those in NTVPM.

34.2.2.1 Basic Approach to the Development of the Model

As noted previously, the model RTVPM treats the track as a system of rigid links connected with frictionless pins, as shown in Figure 34.6. The road wheels, sprockets, and supporting rollers are assumed to be rigidly attached to the vehicle frame. The center of the front idler is, however, considered to be mounted on a pre-compressed spring.

In the analysis, the track system is divided into four sections: the upper run of the track supported by rollers, the lower run of the track in contact with the road wheels and the terrain, the section in contact with the idler, and the section in contact with the sprocket. By considering the equilibrium of various sections of the track system, the interaction between the lower run of the track and the terrain, and the boundary conditions for various track sections, a set of equations can be formulated. The solution to this set of equations determines the track sinkage, inclination of the track system, normal and shear stress distributions on the track–terrain interface, and the track motion resistance, thrust, drawbar pull, and tractive efficiency of the vehicle as functions of track slip. Figure 34.7

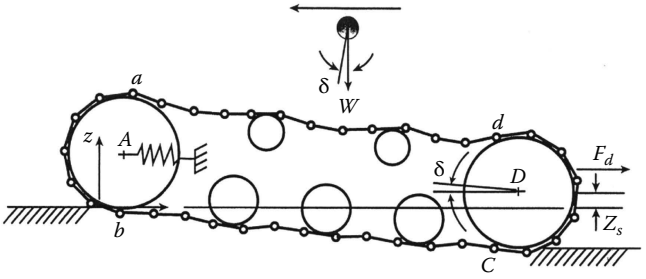


FIGURE 34.6 A schematic of a track system with long-pitch rigid links. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

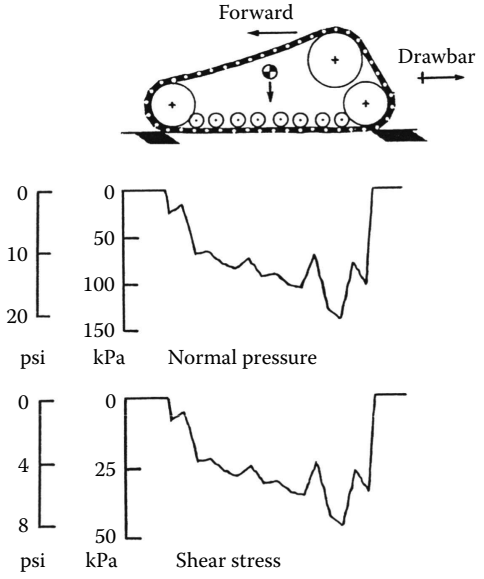


FIGURE 34.7 Normal and shear stress distributions under a tracked vehicle predicted using the computer simulation model RTVPM on clayey soil. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

shows the normal and shear stress distributions under a tracked vehicle with eight road wheels on clayey soil predicted using RTVPM [1,2].

34.2.2.2 Experimental Substantiation

The basic features of RTVPM have been validated with available field test data. Figure 34.8 shows a comparison of the measured and predicted drawbar pull coefficient (drawbar pull to weight ratio) for a heavy tracked vehicle used in construction industry. The terrain was a dry, disked, sandy loam. The measured data shown in the figure were provided by Caterpillar Inc., Peoria, Illinois, United States.

It can be seen that the drawbar pull coefficient of the vehicle predicted using RTVPM agrees reasonably well with the measured. This suggests that the model is capable of providing realistic prediction of the performance of vehicles with long-pitch link tracks in the field.

34.2.2.3 Applications

RTVPM can be used to evaluate the effects of vehicle design and terrain conditions on vehicle performance. It can also be employed in parametric analysis and design optimization of vehicles with rigid link tracks. The applications of RTVPM to design evaluation may be demonstrated through an example for determining the optimum ratio of road wheel spacing to track pitch.

To evaluate the effects of the ratio of road wheel spacing to track pitch, the tractive performances of three track systems were predicted using RTVPM. The three systems have the same track contact length but with five, seven, and eight road wheels and with tracks of various pitches. It has been found that for given overall dimensions of a track system, the ratio of road wheel spacing to track pitch is one of the most significant design parameters that affects the performance of a tracked vehicle with rigid link tracks. Figure 34.9 shows the variation of the drawbar pull coefficient at 20% slip with the ratio of road wheel spacing to track pitch on clayey soil. It is shown that as long as the ratio of road wheel spacing to track pitch is similar, the tractive performances of the three track systems with different numbers of road wheels ranging from five to eight will be similar. It can be seen that on the clayey soil, the optimum ratio of road wheel spacing to track pitch is approximately one. This implies that for a given track pitch, the vehicle designer should select an appropriate road wheel spacing in order to

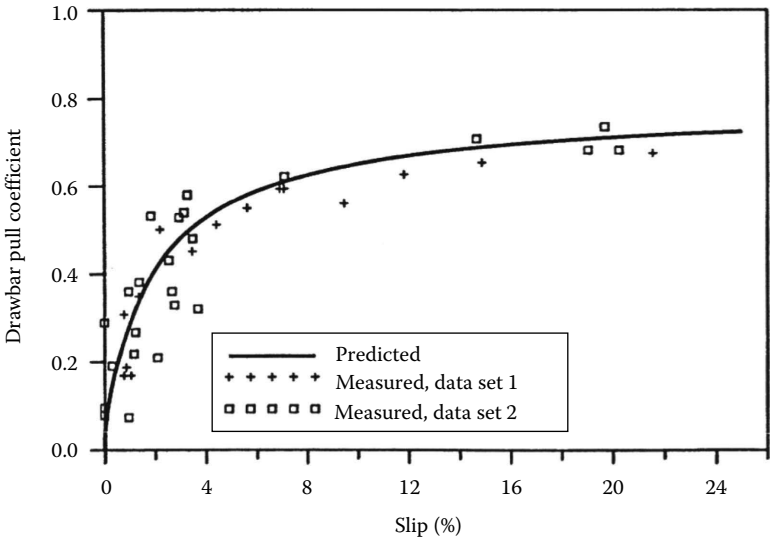


FIGURE 34.8 Comparison of the measured and predicted drawbar pull coefficient of a tracked vehicle on a dry, disked, sandy loam using the computer simulation model RTVPM. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

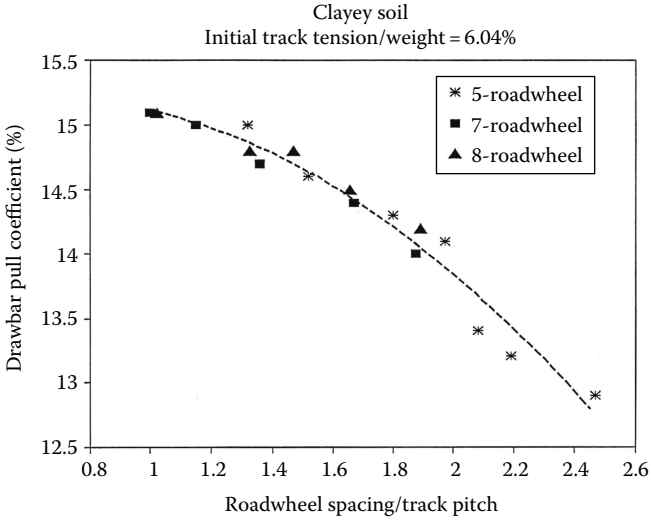


FIGURE 34.9 Effects of the ratio of road wheel spacing to track pitch on drawbar pull coefficient of track systems with different numbers of road wheels on clayey soil. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

achieve the optimum performance. It should be pointed out that track pitch also affects the fluctuation of vehicle forward speed, because of the polygon (or chordal) effect of the sprocket tooth–track link engagement. It can be shown that vehicle speed fluctuation due to the polygon effect is given by

$$\delta_s = 1 - \sqrt{1 - \left(\frac{P}{D}\right)^2} \tag{34.7}$$

where

δ_s is vehicle forward speed fluctuation
P is track pitch
D is the pitch diameter of the sprocket

For most agricultural and industrial tracked vehicles currently in use, the ratio of sprocket pitch diameter to track pitch varies from 3.7 to 4.3. Therefore, vehicle forward speed fluctuation will be in the range of 2.75%–3.72%. If the speed variation is to be kept to 2.75% and if the sprocket pitch diameter is 0.928 m, then the track pitch should be 0.216 m. As shown in Figure 34.9, to ensure good tractive performance on the clayey soil, the ratio of road wheel spacing to track pitch should be approximately one. Therefore, in this case, the optimum road wheel spacing should be approximately 0.216 m.

The example given earlier illustrates a practical application of RTVPM to the selection of optimum design parameters of vehicles with rigid link tracks. RTVPM has been successfully employed in assisting vehicle manufacturers in the development of new products.

34.2.3 Computer Simulation Model NWVPM for Off-Road Wheeled Vehicles

The computer simulation model NWVPM has been developed for performance and design evaluation of off-road wheeled vehicles, including military wheeled vehicles, tractors used in farming and construction industry, and cross-country transport wheeled vehicles. It can also be used in the selection of appropriate tires for a given mission and operating environment. The model takes into

account all major design features of the vehicle, including vehicle weight, axle load distribution, axle suspension stiffness, function of the axle (driven or nondriven), tread of the axle (distance between the left- and right-hand side tires), dynamic axle load transfer, multipass effects of tires (succeeding tires running in the rut formed by the preceding tires), wheelbase, location of the center of gravity, and location of the drawbar hitch. It also takes into consideration all major design features of the tire, including tire diameter, section width, section height, ratio of lug area to carcass area, lug height, inflation pressure and the associated average contact pressure on hard ground, and tire construction (radial or bias).

NWVPM can be used to predict the normal and shear stress distributions on the wheel–terrain interface, the internal (due to hysteresis of tire materials) and external motion resistance of the tire, thrust, drawbar pull, and tractive efficiency of the vehicle as functions of tire slip under steady-state operating conditions. Terrain parameters used in this model are the same as those in NTVPM or RTVPM.

34.2.3.1 Basic Approach to the Development of the Model

NWVPM first predicts the operating mode of the tires (rigid or elastic) on various axles, using the method outlined in Section 20.7.2. The normal pressure distribution on the tire–terrain interface is then evaluated, taking into account the pressure–sinkage characteristics and the response to repetitive loading of the terrain. Based on the normal pressure distribution, the shear stress distribution on the tire–terrain interface at a given tire slip is predicted, following the approach outlined in Section 20.7. The shear stress–shear displacement relationship and response to repetitive shear loading of the terrain are taken into account.

After the normal and shear stress distributions on the tire–terrain interface at a given tire slip have been determined, the motion resistance, thrust, drawbar pull, and tractive efficiency of the wheeled vehicle as functions of slip can be predicted, following the approach outlined in Section 20.7. The dynamic axle load transfer due to drawbar load has been taken into consideration.

34.2.3.2 Experimental Substantiation

The basic features of NWVPM have been substantiated by available field data. Figure 34.10 shows a comparison between the measured and predicted drawbar performance of a tractor on a stubble field [1,2]. The terrain data obtained in the stubble field vary within a certain range. The two solid lines in Figure 34.10 represent the predicted drawbar performance using the upper and lower bounds of the terrain data.

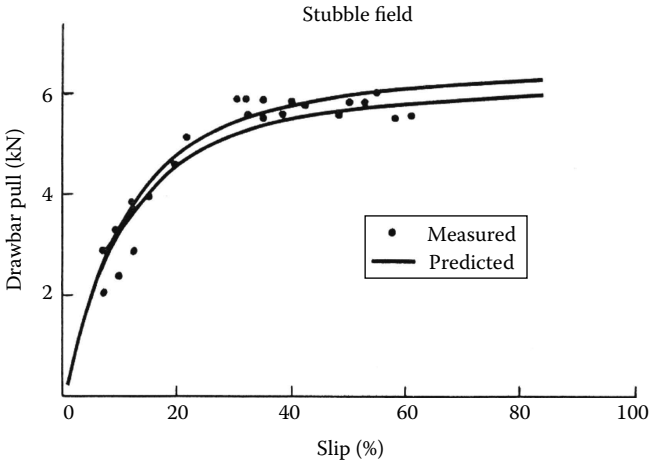


FIGURE 34.10 Comparison of the measured performance of a tractor on a stubble field with that predicted using the computer simulation model NWVPM. (Reproduced from Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.)

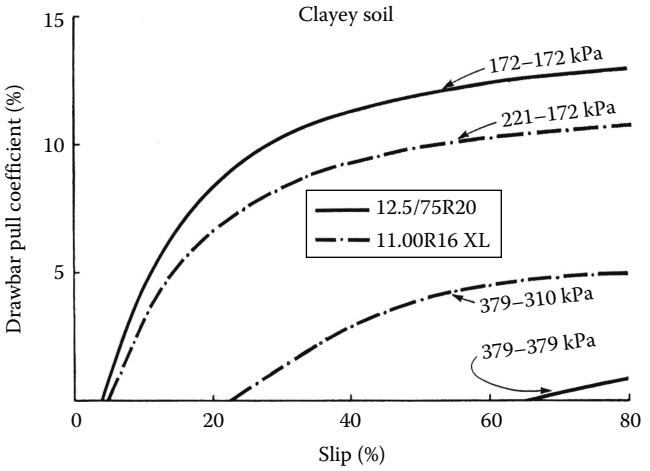


FIGURE 34.11 Drawbar performance of a two-axle, all-wheel-drive vehicle with different tires at various inflation pressure combinations on clayey soil predicted using the computer simulation model NWVPM. (Reproduced from Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.)

It can be seen that the drawbar performance of the tractor predicted using NWVPM agrees reasonably well with the measured data. This suggests that the model is capable of providing realistic prediction of drawbar performance of wheeled vehicles in the field.

34.2.3.3 Applications

NWVPM can be used to evaluate the effects of vehicle and tire design parameters and terrain characteristics on vehicle performance. It can also be employed to select appropriate tires for a given operating environment.

Figure 34.11 shows a comparison of the drawbar performances of a two-axle wheeled vehicle with two types of tire and at different inflation pressure combinations of the front and rear tires over clayey soil, predicted using NWVPM [2]. It is shown that both the tire type and inflation pressure have significant influence on the performance of the two-axle vehicle. For instance, the vehicle with the 12.5/75R20 tire at an inflation pressure of 379 kPa for both the front and rear tires is essentially immobile, whereas at an inflation pressure of 172 kPa for both tires, the performance of the vehicle improves significantly. It can also be seen that the vehicle with the 12.5/75R20 tire at an inflation pressure of 172 kPa for both the front and rear tires has better performance than that with the 11.00R16XL tire with an inflation pressure of 221 kPa at the front and of 172 kPa at the rear (this combination of inflation pressures for the front and rear tires is based on the manufacturer's recommendations). This demonstrates that the model NWVPM can play a useful role in performance and design evaluation of off-road wheeled vehicles and in the selection of appropriate tires for a given vehicle operating over a specific terrain.

In summary, the computer simulation models described earlier are based on understanding of the behavior of the terrain and the mechanics of vehicle–terrain interaction. Consequently, they offer considerable improvements over the empirical and semiempirical methods developed earlier in predicting and evaluating the performance of off-road vehicles. These models are gaining increasingly wide acceptance in industry and governmental agencies in many countries.

34.3 HANDLING CHARACTERISTICS OF OFF-ROAD VEHICLES

The handling characteristics of an off-road vehicle refer to its response to steering command on unprepared terrain. The handling of tracked vehicles is quite different from that of wheeled vehicles. Therefore, they will be discussed separately.

34.3.1 Steering of Tracked Vehicles

There are two basic methods for directional control of a tracked vehicle: one is commonly known as skid-steering and the other is articulated steering. In skid-steering, the thrust on one track (the outside track) is increased and that of the other (the inside track) is reduced or even braked, so as to create a turning moment to overcome the moment of turning resistance due to the skidding of tracks on the ground and the rotational inertia of the vehicle in yaw, as shown in Figure 34.12. Since the moment of turning resistance of the tracks is usually considerable, braking of the inside track is often required. This results in a reduction of the resultant forward thrust. On soft terrain, this often leads to immobilization. For tracked vehicles with two or more units (or chassis), steering may be accomplished by rotating one unit against the other using a steering joint to make the vehicle follow a prescribed curved path (i.e., articulated steering), as shown in Figure 34.13. In articulated steering,

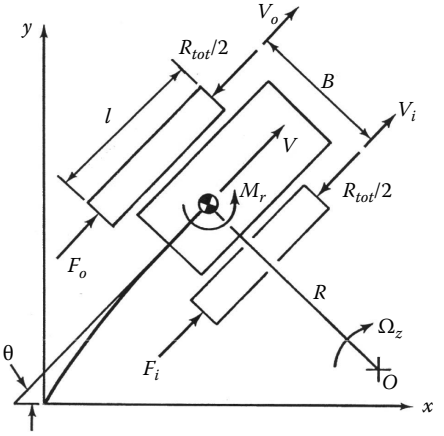


FIGURE 34.12 Skid-steering for tracked vehicles.

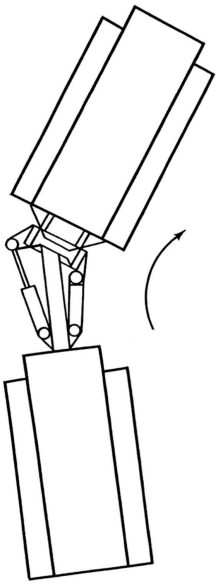


FIGURE 34.13 Articulated steering for tracked vehicles. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

turning is initiated by activating the steering joint between the two units, and no adjustment of thrusts of the outside and inside tracks is required. Thus, the resultant forward thrust of the vehicle can be maintained during a turn. On a soft terrain, articulated steering can provide tracked vehicles with better mobility during turning maneuvers than skid-steering.

34.3.2 Simplified Analysis of the Mechanics of Skid-Steering for Tracked Vehicles

The turning behavior of a tracked vehicle using skid-steering depends on the thrusts of the outside and inside tracks F_o and F_i, the resultant motion resistance R_{tot}, the moment of turning resistance M_r exerted on the track by the ground, and the vehicle dimensions, as shown in Figure 34.12. For low-speed turning maneuvers, the effects of centrifugal force may be neglected and the behavior of the vehicle can be described by the following two equations of motion:

$$m\frac{d^2s}{dt^2} = F_o + F_i - R_{tot} \tag{34.8}$$

$$I_z = \frac{d^2\theta}{dt^2} = \frac{B}{2}(F_o - F_i) - M_r \tag{34.9}$$

where

s is the displacement of the center of gravity of the vehicle
θ is the angular displacement of the vehicle
B is the tread of the vehicle (i.e., the distance between the centerlines of the two tracks)
I_z and m are the mass moment of inertia about the vertical axis passing through its center of gravity and the mass of the vehicle, respectively

With known initial conditions, the earlier two differential equations can be integrated with respect to time, and the path of the center of gravity and the orientation of the vehicle can be determined as functions of time.

Under steady-state conditions, both linear and angular accelerations are zero. Equations 34.8 and 34.9 become

$$F_o + F_i - R_{tot} = 0 \tag{34.10}$$

$$\frac{B}{2}(F_o - F_i) - M_r = 0 \tag{34.11}$$

The thrusts of the outside and inside tracks required to maintain a steady-state turn are therefore expressed by

$$F_o = \frac{R_{tot}}{2} + \frac{M_r}{B} = \frac{f_r W}{2} + \frac{M_r}{B} \tag{34.12}$$

$$F_i = \frac{R_{tot}}{2} - \frac{M_r}{B} = \frac{f_r W}{2} - \frac{M_r}{B} \tag{34.13}$$

where

f_r is the coefficient of motion resistance of the vehicle in the longitudinal direction
W is the vehicle weight

To determine the thrusts F_o and F_i, the moment of turning resistance M_r must be known. It can be determined experimentally or analytically. If the normal pressure is uniformly distributed along the track contact length, the lateral resistance per unit track length R_l be expressed by

$$R_l = \frac{\mu_t W}{2l} \tag{34.14}$$

where

μ_t is the coefficient of lateral resistance (friction)
l is the contact length of each of the two tracks

The value of μ_t depends not only on the terrain but also on the design of the track. Over soft terrain, the vehicle sinks into the ground, and the tracks together with the grousers will be sliding on the terrain, as well as displacing the soil laterally during turning maneuvers. The lateral forces acting on the tracks and the grousers due to displacing the soil laterally form part of the lateral resistance. Table 34.1 shows the average values of μ_t for steel and rubber tracks on various types of surface [1,13].

Assuming that the coefficient of lateral resistance μ_t is a constant, the resultant moment of the lateral resistance about the centers of the two tracks M_r (i.e., the moment of turning resistance of the vehicle) can be expressed by

$$M_r = 4\frac{\mu_t W}{2l}\int_o^{1/2} x dx = \frac{\mu_t W l}{4} \tag{34.15}$$

It should be emphasized that the value of M_r as calculated by Equations 34.15 is for a vehicle with uniform contact pressure and that the value of μ_t is a constant. Experimental evidence shows that in practice, the values of the moment of turning resistance M_r and the corresponding coefficient of lateral resistance μ_t appear to be functions of turning radius. This issue will be discussed further later.

Making use of Equation 34.15, Equations 34.12 and 34.13 can be rewritten in the following form:

$$F_o = \frac{f_r W}{2} + \frac{\mu_t W l}{4B} \tag{34.16}$$

$$F_i = \frac{f_r W}{2} - \frac{\mu_t W l}{4B} \tag{34.17}$$

TABLE 34.1
Values of Lateral Resistance Coefficient of Various Types of Track

	Coefficient of Lateral Resistance μ_t		
Track Material	**Concrete**	**Hard Ground (Unpaved)**	**Grass**
Steel	0.50–0.51	0.55–0.58	0.87–1.11
Rubber	0.90–0.91	0.65–0.66	0.67–1.14

Source: Hayashi, I., Practical analysis of tracked vehicle steering depending on longitudinal track slippage, *Proceedings of the 5th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. II, Detroit, MI, 1975.

Equations 34.16 and 34.17 are useful in estimating the steerability of a tracked vehicle. As discussed in Chapter 20, the maximum thrust of a track is limited by the shear strength of the terrain, track contact area, and normal load on the track (see Equation 20.9 in that chapter). Accordingly, the maximum thrust that can be developed by the outside track is given by

$$F_{omax} \le cbl + \frac{W \tan \phi}{2} \tag{34.18}$$

where
b is the track width
l is the track contact length
c and ϕ are the cohesion and angle of internal shearing resistance of the terrain

Substituting Equation 34.16 into Equation 34.18, one obtains

$$\frac{f_r W}{2} + \frac{\mu_t W l}{4B} \le cbl + \frac{W \tan \phi}{2}$$

or

$$\frac{l}{B} \le \frac{1}{\mu_t}\left(\frac{4cA}{W} + 2\tan\phi - 2f_r\right) \tag{34.19}$$

where A is the contact area of one track.

This indicates that to steer a tracked vehicle without spinning the outside track, the ratio of track length to tread of the vehicle l/B must satisfy the following condition:

$$\frac{l}{B} \le \frac{2}{\mu_t}\left(\frac{c}{p} + \tan\phi - f_r\right) \tag{34.20}$$

where p is the average normal pressure on the track, which is equal to $W/2A$.

On a sandy terrain with $c = 0$, $\phi = 30°$, $\mu_t = 0.5$, and $f_r = 0.1$, the value of l/B should be less than 1.9. In other words, if the ratio of contact length to tread of a tracked vehicle is greater than 1.9, the vehicle will be unable to steer on the sandy terrain. On a clayey terrain, with $c = 3.45$ kPa, $\phi = 10°$, $p = 6.9$ kPa, $\mu_t = 0.4$, and $f_r = 0.1$, the value of l/B must be less than 2.88. These examples show the importance of the ratio of contact length to tread l/B to the steerability of a tracked vehicle.

34.3.3 General Theory for Skid-Steering

The analysis of the mechanics of skid-steering described in the preceding section is based on a number of simplifying assumptions. Consequently, the method can only provide preliminary prediction of the turning behavior of tracked vehicles. In an attempt to improve the prediction of the handling characteristics of tracked vehicles, Steeds developed a more detailed analysis of the mechanics of skid-steering [14]. In his analysis, the shear stress developed on the track–ground

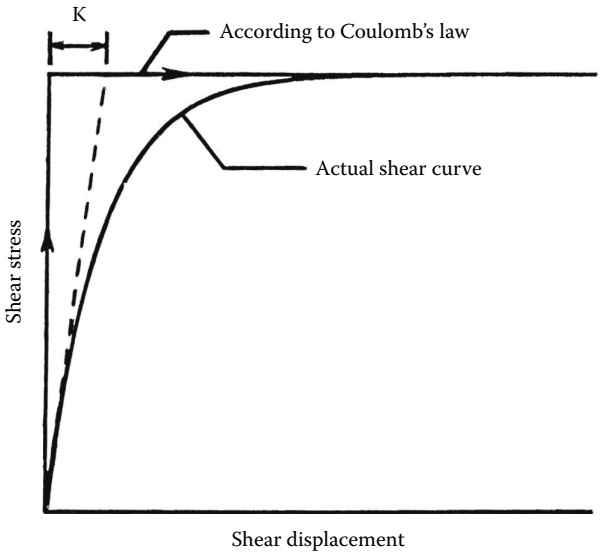


FIGURE 34.14 Comparison between the shear stress–shear displacement relationship measured in the field and that based on Coulomb's law of friction. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

interface is assumed to obey the Coulomb law of friction. The friction may be considered to be either isotropic or anisotropic. In the latter case, different values of coefficient of friction are assigned to the longitudinal and lateral directions of the track. It should be noted, however, that the Coulomb law of friction implies that as soon as a small relative movement between a track element and the ground is initiated, the maximum shear stress will develop instantly on that track element. Experimental evidence presented in Chapter 20 indicates that the shear stress will reach its maximum value only after a certain shear displacement has taken place, as shown in Figure 34.14. Field test data show that considerable discrepancy exists between the steering behavior of tracked vehicles predicted using Steeds' method and that measured. An improved, general theory for the mechanics of skid-steering on firm ground has, therefore, been developed recently by Wong and Chiang [1,15].

34.3.3.1 Basic Approach to the Development of the General Theory

The improved, general theory of skid-steering for tracked vehicles is based on the following premises [1,15]:

1. The ground is firm. Consequently, track sinkage and associated bulldozing effect of the track in the lateral direction during a turning maneuver are neglected.
2. The shear stress developed at a given point on the track–ground interface during a turn is dependent upon the shear displacement at that point, measured from its initial contact with the ground. For rubber-belt tracks or steel tracks with rubber pads, the shear stress is that developed between the rubber and the ground.
3. The direction of the shear stress at a point on the track–ground interface is opposite to that of the sliding velocity between the track and the ground at that point.
4. The component of the shear stress along the longitudinal direction of the track constitutes the tractive or braking effort, while the lateral component forms the lateral resistance of the track. The moment of the lateral resistance about the turn center of the track constitutes the moment of turning resistance.

The procedure for predicting the turning behavior of tracked vehicles using the general theory is briefly outlined in the following, and the details may be found in references [1,15]:

a. For a specific turning maneuver at a given vehicle forward speed and turning radius, the shear displacements of the contact points on the track–ground interface are first calculated, based on a kinematic analysis.
b. With known shear stress–shear displacement relationship, the distribution of shear stress on the track–ground interface is determined.
c. Based on the direction of the sliding velocity of the contact point relative to the ground, the direction of the shear stress at each contact point is evaluated.
d. The shear stress at each contact point is resolved into a longitudinal and a lateral component with respect to the track.
e. The integration of the longitudinal component of shear stress over the entire track contact area gives the resultant longitudinal force of the track, and the corresponding sprocket torque is equal to the product of the longitudinal force and sprocket pitch radius.
f. The integration of the lateral component of the shear stress over the entire track contact area gives the resultant lateral track force. The moment of the resultant lateral force about the turn center of the track determines the moment of turning resistance of the track.
g. From the analysis noted earlier, the dynamic equilibrium equations for a tracked vehicle in a steady-state turning maneuver are as follows (see Figure 34.15):

$$F_{xo} + F_{xi} = \frac{WV^2}{gR'}\cos\beta \tag{34.21}$$

$$F_{yo} + F_{yi} = \frac{WV^2}{gR'}\sin\beta + (R_o + R_i) \tag{34.22}$$

$$M_{Lo} - M_{Li} - \frac{B}{2}(R_o - R_i) + (s_o\cos\beta + c_x\sin\beta)\frac{WV^2}{gR'} = M_{ro} + M_{ri} \tag{34.23}$$

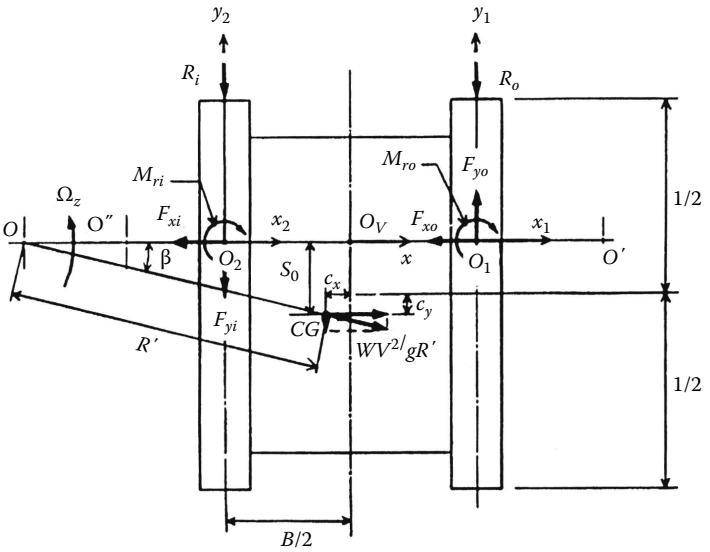


FIGURE 34.15 Forces and moments acting on a tracked vehicle during a steady-state turn. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

where

F_{xo} and F_{xi} are the resultant lateral forces acting on the outside and inside tracks, respectively
F_{yo} and F_{yi} are the resultant longitudinal forces acting on the outside and inside tracks, respectively
M_{Lo} and M_{Li} are the turning moments of the resultant longitudinal forces about the center O_v (shown in Figure 34.15) of the outside and inside tracks, respectively
M_{ro} and M_{ri} are the moments of turning resistance acting on the outside and inside tracks, respectively
R_o and R_i are the longitudinal motion resistances acting on the outside and inside tracks, respectively
g is the acceleration due to gravity
R' is the turning radius
s_o is the longitudinal offset of the turning center with respect to the center of gravity of the vehicle, caused by the centrifugal force
V is the vehicle forward speed
W is the vehicle weight

All other parameters are shown in Figure 34.15.

As noted previously, the shear stress on the vehicle–ground interface is related to the kinematics of the vehicle (i.e., vehicle forward speed V, turning radius R', and longitudinal offset s_o). Consequently, the forces and moments developed on the track–ground interface are functions of ω_o, ω_i, and s_o. For a specific turning maneuver with a given vehicle forward speed and turning radius, the values of ω_o, ω_i, and s_o can be determined by solving Equations 34.21 through 34.23 simultaneously. Thus, all forces and moments acting on the vehicle during a given steady-state turning maneuver are completely defined.

Various normal pressure distributions under the track can be accommodated in the general theory. They include normal loads on the tracks being concentrated beneath the road wheels, normal loads supported only by the track links immediately under the road wheels, and normal loads distributed continuously over the track contact length.

34.3.3.2 Experimental Substantiation

To validate the basic features of the general theory described earlier, the steering behavior of a military tracked vehicle was predicted, and the results were compared with available measured data reported in Reference [16]. Figure 34.16 shows the variations of the predicted sprocket torques for the outside and inside tracks with the theoretical turning radius (derived from the angular speeds of the outside and inside track sprockets ω_o, ω_i, neglecting the slip or skid of the tracks) at various vehicle forward speeds with different types of normal load distribution. In the figure, the measured data reported in reference [16] are also shown. For comparison, the predicted results obtained using Steeds' theory are presented as well.

It can be seen that predictions obtained using the general theory bear a strong resemblance to the measured data. Both show that the magnitudes of the sprocket torques generally decrease with the increase of theoretical turning radius. In contrast, predictions obtained using Steeds' theory differ greatly from the measured data. The sprocket torques predicted using Steeds' method remain essentially constant over a wide range of turning radii, particularly at low speeds. The reason for the poor predictions using Steeds' theory is primarily due to the assumption that the shear stress on the track–ground interface reaches its maximum value instantly, as soon as a small relative movement between the track and the ground takes place. Figure 34.17a and b show the lateral shear stress distributions along the longitudinal centerline of the outside and inside tracks, respectively, at various turning radii. It is noted that even at the front contact point where a track element just comes into contact with the ground, the lateral stress is at its maximum. In contrast, based on the general theory, the lateral stress at the front contact point of the outside and inside tracks is zero, as shown in Figure 34.18a and b, respectively. It should also be pointed out that the lateral shear

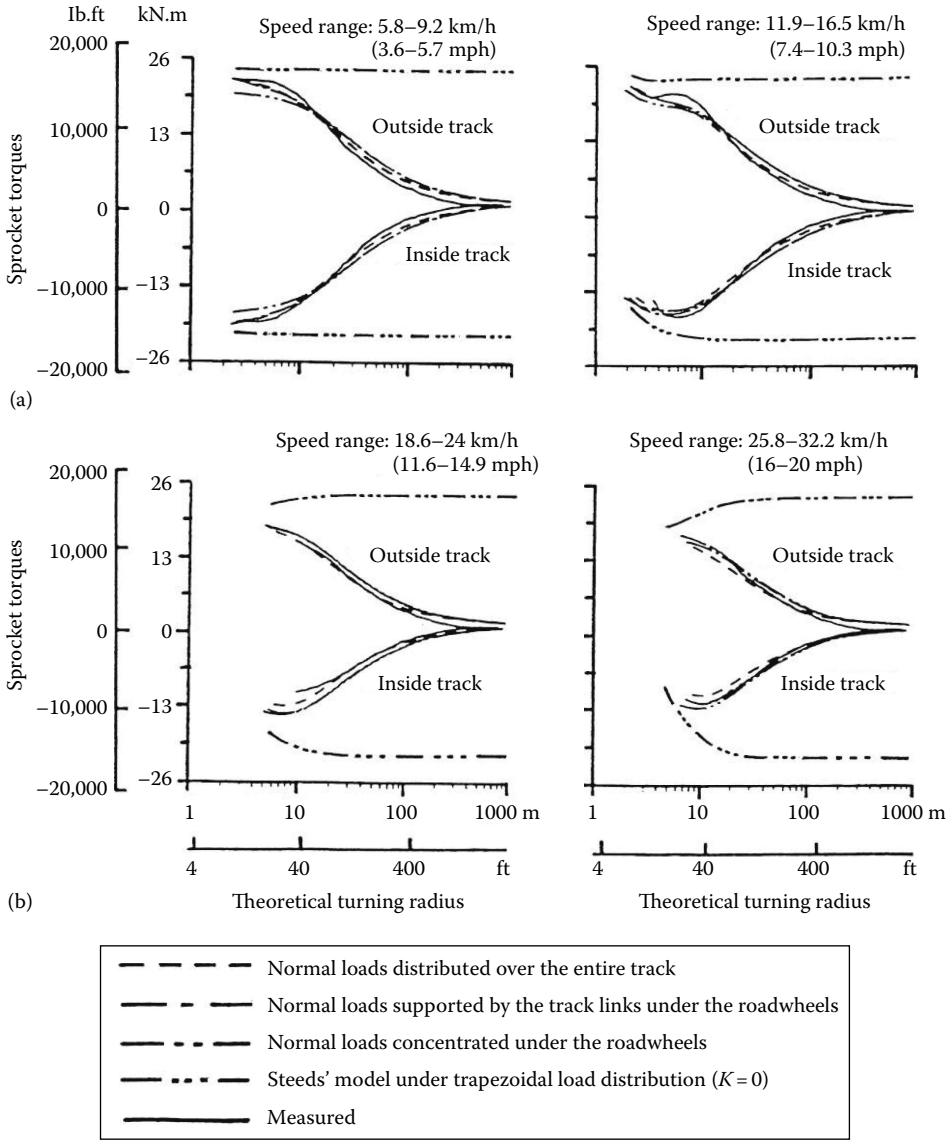


FIGURE 34.16 Comparison of the measured and predicted relationships between sprocket torques and theoretical turning radius for a tracked vehicle at various forward speeds using the general theory. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

stress distributions on the outside and inside tracks predicted using Steeds' theory remain essentially the same when the turning radius is larger than a certain value, such as 50 m shown in Figure 34.17. Since the moment of the lateral shear force about the turn center of the track constitutes the moment of turning resistance, this leads to essentially constant moments of turning resistance for the outside and inside tracks over a wide range of turning radii, based on Steeds' theory.

During a steady-state turn, the sprocket torques at the outside and inside tracks are to generate the required tractive or braking forces, which form a turning moment for the tracked vehicle to overcome primarily the resultant moment of turning resistance caused by the lateral skidding of the tracks on the ground. Consequently, as shown in Figure 34.16, based on Steeds' method, the sprocket torques for the outside and inside tracks remain essentially constant when the turning radius is larger than a certain value. In contrast, based on the general theory, the

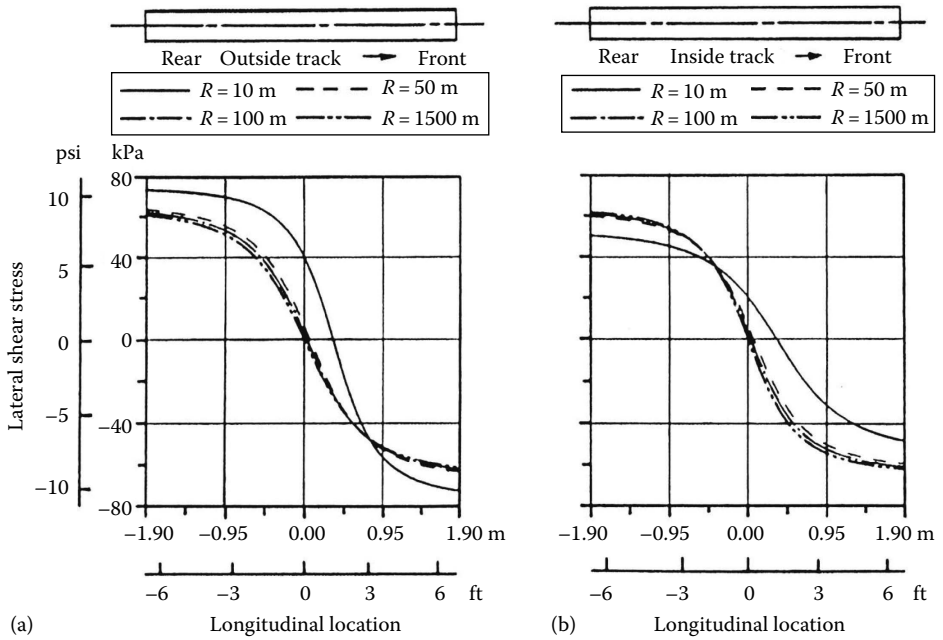


FIGURE 34.17 Lateral shear stress distributions along the centerline of (a) the outside track and (b) the inside track at different turning radii according to Steeds' theory. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

magnitude of the lateral shear stress decreases with the increase of turning radius, as shown in Figure 34.18. As a result, the magnitudes of the sprocket torques for the outside and inside tracks decrease with the increase of turning radius, as shown in Figure 34.16. This shows that the general theory provides much more realistic and accurate predictions of steering behavior of tracked vehicles than Steeds' theory. It is also interesting to note from Figure 34.16 that on firm ground the normal load distribution does not have a significant influence on the turning behavior of tracked vehicles.

34.3.3.3 Coefficient of Lateral Resistance

In the simplified analysis of the mechanics of skid-steering given in Section 34.3.2, a coefficient of lateral resistance μ_t (also known as the coefficient of lateral friction) is introduced in predicting the lateral force and moment of turning resistance acting on a track during a turning maneuver. If normal pressure is uniformly distributed along the track, the moment of turning resistance M_r due to track skidding is given by Equation 34.15.

Equation 34.15 indicates that the turning resistance moment M_r is independent of turning radius. Experimental evidence shows, however, that in practice the moment of turning resistance M_r varies with turning radius. As an expedient, it has been suggested that the coefficient of lateral resistance μ_t in Equation 34.15 be expressed as a function of turning radius. A number of empirical equations have been proposed to correlate μ_t with turning radius. These empirical equations contain a number of empirical constants and there is no evidence that these empirical relations can be generally applied.

As described previously, using the general theory, the moment of turning resistance M_r can be predicted analytically, using only the coefficient of friction μ that defines the shearing property of the track–ground interface. This means that using the general theory, it is not necessary to introduce a coefficient of lateral resistance μ_t in predicting the moment of turning resistance M_r. In fact the corresponding coefficient of lateral resistance μ_t used in Equation 34.15 can be derived quantitatively by equating the moment of turning resistance calculated from the general theory to that given by Equation 34.15. As the moment of turning resistance M_r calculated

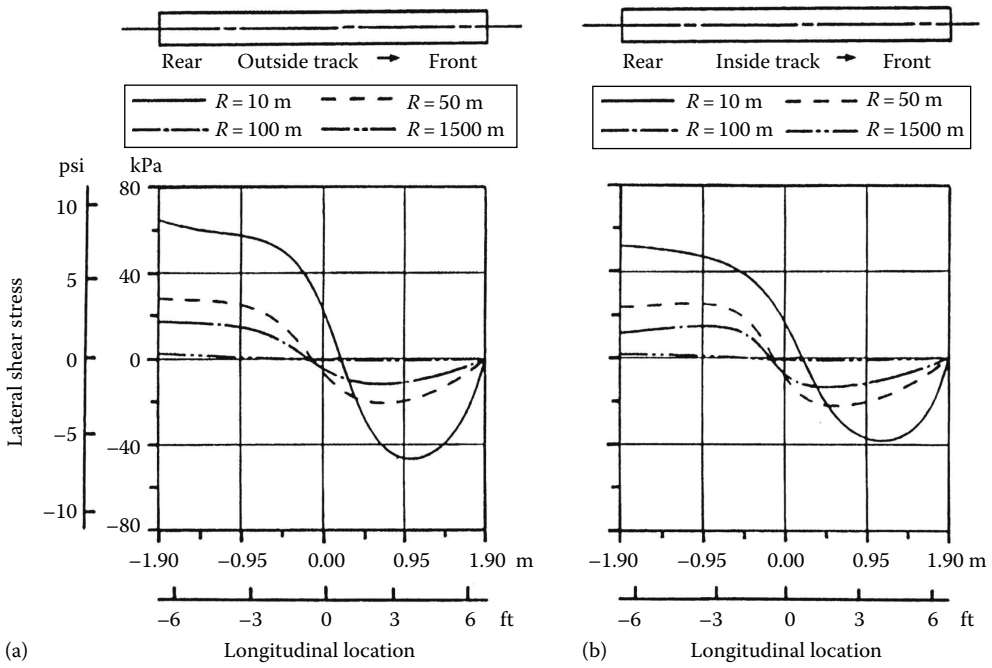


FIGURE 34.18 Lateral shear stress distributions along the centerline of (a) the outside track and (b) the inside track at different turning radii according to the general theory. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

by the general theory is related to turning radius, the relationship between the corresponding coefficient of lateral resistance μ_t used in Equation 34.15 and turning radius can be established.

As an example of the application of this method, Figure 34.19 shows the variations of the corresponding coefficient of lateral resistance μ_t with theoretical turning radius for the military tracked vehicle at a speed of 7.5 km/h, under different ground conditions, with the value of coefficient of

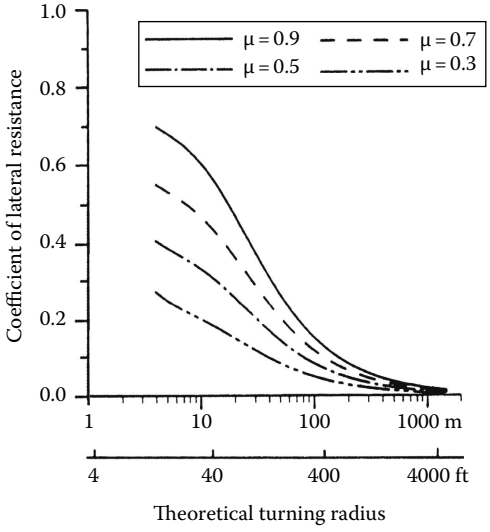


FIGURE 34.19 Variations of the coefficient of lateral resistance with turning radius predicted using the general theory for a tracked vehicle on surfaces with different values of coefficient of friction. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

friction μ varying from 0.3 to 0.9. This demonstrates that if Equation 34.15 is used to predict the moment of turning resistance, the relationship between μ_t and turning radius can be derived from the general theory for a given ground condition with known value of coefficient of friction μ. This means that by applying the general theory, the time-consuming field tests for defining the relation between μ_t and turning radii can be completely eliminated.

In summary, it appears that the general theory presented earlier provides an improved method for analyzing the mechanics of skid-steering and for predicting the turning behavior of tracked vehicles on firm ground.

34.3.4 Steering of Off-Road Wheeled Vehicles

Steering of most off-road wheeled vehicles is effected by changing the heading of front wheels through a steering system, similar to that of road vehicles. For multiaxle off-road wheeled vehicles, such as armored vehicles, in addition to the front wheels, intermediate wheels are in some cases made steerable. The prime consideration in the design of steering system for wheeled vehicles is minimum tire scrub during cornering. This requires that during the turn, all tires should be in pure rolling with minimum lateral sliding. To satisfy this requirement, the wheels should follow curved paths with different radii originating from a common turn center, as shown in Figure 34.20, for a 6 × 6 vehicle with steerable front and intermediate wheels. This establishes the proper relationship between the steer angle of the inside wheel and that of the outside wheel on the same axle. From Figure 34.20, it can be seen that for the 6 × 6 vehicle, the steer angles of outside and inside front wheels δ_{o1} and δ_{i1} and those of the outside and inside intermediate wheels δ_{o2} and δ_{i2} should satisfy the following relationships, respectively:

$$\cot\delta_{o1} - \cot\delta_{i1} = \frac{B}{L_1}$$
$$\cot\delta_{o2} - \cot\delta_{i2} = \frac{B}{L_2} \tag{34.24}$$

where

B is the tread (or track)

L_1 and L_2 are wheelbases, as shown in Figure 34.20

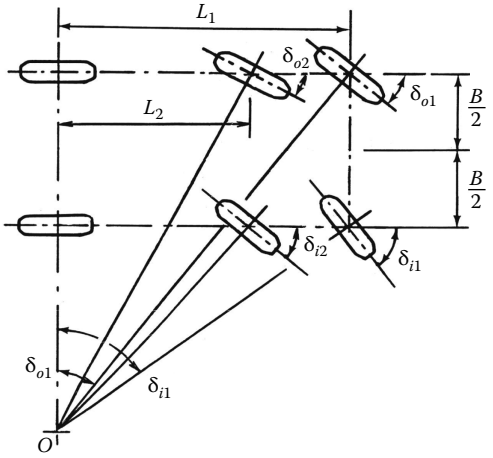


FIGURE 34.20 Desirable steering geometry for wheeled vehicles with steerable front and intermediate wheels.

The steering geometry that satisfies Equation 34.24 is usually referred to as the Ackermann steering geometry.

In many respects, the handling of off-road wheeled vehicles is similar to that of road vehicles. Since the handling characteristics of road vehicles are covered extensively in other chapters of this handbook, their discussion is not repeated here. However, there are differences between them. For instance, in many cases, the speed of off-road vehicles operating on unprepared terrain is much lower than that of road vehicles running on paved highways. Consequently, for these cases, the effects of centrifugal force may be neglected in the analysis of the handling of off-road wheeled vehicles. Furthermore, on unprepared terrain, the side (cornering) force–slip angle relationships of tires are usually quite different from those on paved road. Figure 34.21 shows the variations of the side (cornering) force coefficient (i.e., the ratio of side [cornering] force to normal load) with slip angle of a tractor front tire (7.5-18AS) on road and on a variety of unprepared terrains [17]. It can be seen that there are noticeable differences between the cornering behavior of a tire operating on unprepared terrain and that on paved road. It should be mentioned that the cornering behavior of tires on unprepared terrain has not been studied, both analytically and experimentally, as extensively as that of tires on paved road. This is an area that requires further research in order to be able to predict the handling behavior of off-road wheeled vehicles on unprepared terrain more accurately [18].

In addition to the Ackermann steering described earlier, articulated steering is also employed in off-road wheeled vehicles, particularly in heavy off-road wheeled vehicles used in earthmoving and construction industry. Detailed analysis of articulated steering of off-road wheeled vehicles may be found in Reference [19].

In summary, while the simplified analysis of skid-steering provides a basic understanding of the handling behavior of tracked vehicles, it cannot accurately predict the turning characteristics of tracked vehicles. The general theory for skid-steering developed by Wong and Chiang provides an improved method for predicting the turning behavior of tracked vehicles. Its basic features have been substantiated by field test data. For off-road wheeled vehicles, to improve the prediction of their handling behavior, further research into the cornering behavior of tires on unprepared terrain is required.

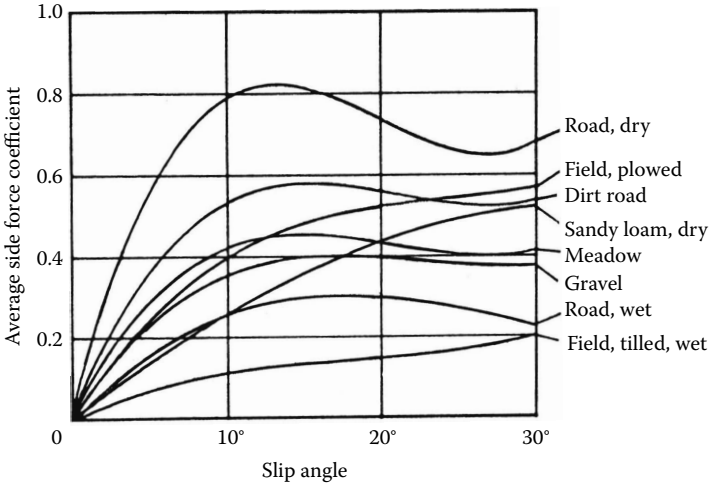


FIGURE 34.21 Comparison of the cornering characteristics of a tractor front tire on various types of terrain with those on paved road. (Reproduced from Schwanghart, H., Measurements of forces on steered, towed wheels (in German), *Proceedings of the 7th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. I, Calgary, Alberta, Canada, 1981. With permission from the International Society for Terrain-Vehicle Systems.)

34.4 RIDE CHARACTERISTICS OF OFF-ROAD VEHICLES

Ride quality is concerned with the vibration of the vehicle excited by surface irregularities and with the effects of vehicle vibration on the comfort of the driver and passenger. To evaluate the ride quality of off-road vehicles, it is essential to have a basic understanding of the human response to vibration, of the characteristics of surface irregularities, and of the modeling of the vibration behavior of the vehicle (see also Chapters 29 and 36).

34.4.1 Ride Comfort Criteria

Over the years, a great deal of research has been conducted to define the limits of acceptable human exposure to whole-body vibration. Among the ride comfort criteria that have been proposed, the International Standard ISO 2631 is the most widely used [1,20]. It takes into account four physical factors that are of primary importance to the human response to vibration, namely, the intensity, the frequency, the direction, and the duration (exposure time) of the vibration. The International Standard 2631-1:1985(E) published in 1985 defines three distinct limits (boundaries) for whole-body vibration in the frequency range of 1–80 Hz:

A. Fatigue-Decreased Proficiency Boundary
This boundary specifies a limit beyond which exposure to vibration can be regarded as carrying a significant risk of impaired working efficiency in many kinds of task, such as vehicle driving.

B. Exposure Limit (Health or Safety)
The exposure limit as a function of frequency and exposure time is of the same general form as the fatigue-decreased proficiency boundary, but the corresponding levels are raised by a factor of 2 (6 dB higher). Exceeding the exposure limit is not recommended without special justification and precautions.

C. Reduced Comfort Boundary
Values of the reduced comfort boundary are obtained by lowering the corresponding values for the fatigue-decreased proficiency boundary by a factor of 3.15 (10 dB lower). In the transport situation, the reduced comfort boundary is related to difficulties in carrying out such activities as eating, reading, and writing.

The latest edition of the International Standard 2631-1:1997(E), published in 1997, recommends the use of frequency-weighted root-mean-square (rms) acceleration for evaluating whole-body vibration for health, comfort, and perception in the frequency range of 0.5–80 Hz and for motion sickness in the frequency range of 0.1–0.5 Hz. In the latest edition, potential effects of vibration on human performance (i.e., the fatigue-decreased proficient boundary in the previous edition) are no longer covered. For vibration with crest factor less than 9 (the crest factor is defined as the modulus of the ratio of the maximum instantaneous peak value of the frequency-weighted acceleration signal to its rms value), Table 34.2 describes the relations between the values of vibration magnitude and indications of comfort [20].

The "absorbed power," which represents the time rate of vibration energy transmitted to the human body, has also been proposed as a parameter of significance in evaluating human tolerance to vibration [1,21]. It is used in evaluating the ride quality of military vehicles negotiating rough terrain, especially in the United States. Currently, the tolerance limit is taken to be 6 W absorbed power at the driver's position. The ride-limiting speed is taken as the speed at which the driver's average absorbed power over the total elapsed time reaches a sustained level of 6 W. For a vibration environment characterized by a high crest factor, peak vertical acceleration is often used in conjunction with absorbed power to assess ride quality. The suggested limit for the peak vertical acceleration on the floor beneath the driver's seat is 2.5 *g*.

TABLE 34.2
Relations between Values of Vibration Magnitude and Indications of Comfort

Acceleration (rms) (m/s^2)	Indications
Less than 0.315	Not uncomfortable
0.315–0.63	A little uncomfortable
0.5–1.0	Fairly uncomfortable
0.8–1.6	Uncomfortable
1.25–2.5	Very uncomfortable
Greater than 2	Extremely uncomfortable

Source: ISO 2631-1:1997(E).

34.4.2 Characterization of Surface Irregularities

To predict vehicle vibration excited by surface irregularities, two methods are usually used to define the surface profile. If the time history of vehicle vibration is of interest, then surface profile is measured and directly used as input to vehicle vibration analysis. This type of analysis is usually referred to as the time domain analysis. The time history of vehicle vibration may be processed to obtain the vibration spectrum, from which the relationship between vibration intensity, such as the rms acceleration, and frequency of excitation may be established. The other method is to regard the surface profile as a random function and to characterize the surface profile in terms of power spectral density as a function of spatial frequency (expressed in terms of cycle/m) [1]. For a linear vehicle system, through its transfer function, vehicle vibration (as output) excited by surface irregularities (as input) can be predicted. This type of analysis is usually referred to as the frequency domain analysis. It has been found that in many cases, the relationship between the power spectral density of surface profile and spatial frequency may be approximated by

$$S_g(\Omega) = C_{sp}\Omega^{-N} \tag{34.25}$$

where

$S_g(\Omega)$ is the power spectral density function of surface profile in m^2/cycle/m
Ω is the spatial frequency in cycle/m
C_{sp} and N are constants for a given surface

Table 34.3 shows the values of C_{sp} and N for various types of surface [1].

In vehicle vibration analysis, it is more convenient to express the power spectral density of surface profile in terms of temporal frequency in Hz, rather than in terms of spatial frequency in cycle/m. The transformation of the spatial frequency Ω in cycle/m to the temporal frequency f in Hz is through the speed of the vehicle V:

$$f(\text{Hz}) = \Omega\ (\text{cycle/m})V(\text{m/s}) \tag{34.26}$$

The transformation of power spectral density of surface profile expressed in terms of spatial frequency $S_g(\Omega)$ to that in terms of temporal frequency $S_g(f)$ is also through the speed of the vehicle V:

$$S_g(f) = \frac{S_g(\Omega)}{V} \tag{34.27}$$

TABLE 34.3
Values of C_{sp} and N for Various Types of Surface

Surface Type	N	C_{sp}
Smooth highway	2.1	4.8×10^{-7}
Highway with gravel	2.1	4.4×10^{-6}
Pasture	1.6	3.0×10^{-4}
Plowed field	1.6	6.5×10^{-4}

Source: Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.

For a linear vehicle system with known transfer function $|H(f)|$, the power spectral density function $S_v(f)$ representing the vibration of the vehicle and the power spectral density function $S_g(f)$ representing the surface profile are related by

$$S_v(f) = |H(f)|^2 S_g(f) \tag{34.28}$$

After the power spectral density function $S_v(f)$ for the vehicle has been determined, the rms acceleration of the vehicle can be determined as a function of temporal frequency f. The results can then be compared with selected ride comfort criteria, such as the International Standard ISO 2631.

It should be pointed out that if the deformation of the terrain under vehicular load is significant, then the terrain profile will be modified as the vehicle traversing it. This will complicate the analysis of the vibration of off-road vehicles. The modification of terrain profile by vehicular load and its effects on vehicle vibration over deformable terrain are topics that so far have not been investigated extensively and warrant further research.

34.4.3 Modeling Ride Dynamics of Off-Road Vehicles

The approach to the modeling of the ride dynamics of off-road wheeled vehicles is similar to that of road vehicles. Since the ride dynamics of road vehicles is extensively covered in other chapters of this handbook, it will not be repeated here.

Modeling the ride dynamics of tracked vehicles, in certain respects, is similar to that of off-road wheeled vehicles, as they have similar suspension elements, such as springs and dampers. However, the presence of the track wrapping around the road wheels, sprockets, idlers, and supporting rollers introduces a great deal of complexity (see Figure 34.22).

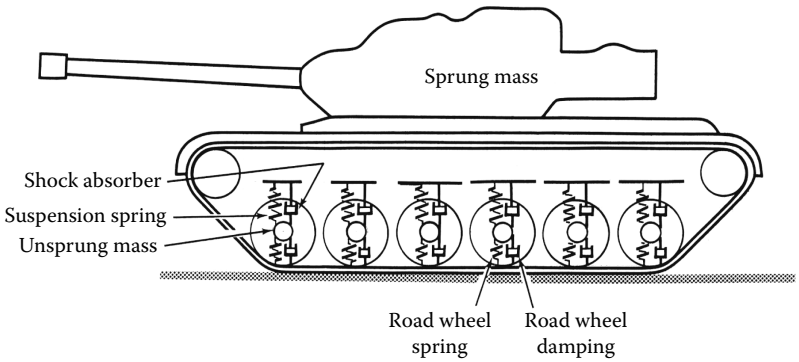


FIGURE 34.22 A ride model for a military tracked vehicle. (Reproduced from Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.)

The track acts as a mechanism to transfer mass between the vehicle's sprung and unsprung parts, as an elastic element with variable stiffness, as a terrain profile filtering mechanism, as a source of additional damping, and as an additional system capable of initiating its own vibrations. In existing ride dynamics models for tracked vehicles, a variety of assumptions are used in modeling the track. In some models, the track is idealized as a continuous flexible belt, inextensible or extensible with known longitudinal stiffness, whereas in others, the track is assumed as a system of rigid links connected with pins or rubber bushings.

Various assumptions on track tension distribution, ranging from constant track tension, through equal step increase (or decrease) in track tension between road wheels, to continuously variable track tension, have been made in existing ride dynamics models for tracked vehicles [22,23]. However, it appears that no existing model has taken into account the shearing action between the track and terrain and the resulting variation of track tension along the track.

In an earlier version of the vehicle dynamics module, VEHDYN, of the NATO Reference Mobility Model (NRMM), the track is modeled in the following way [22,24]:

a. If the terrain profile interferes with the tangent between the sprocket (or idler) and adjacent road wheel, the track will deflect to conform to the profile. This deflection will give rise to an increase in track tension, which will be transmitted through the sprocket (or idler) and road wheel to the vehicle. The increase in track tension is estimated by assuming that a virtual linear spring acts perpendicularly to the tangent between the sprocket (or idler) and adjacent road wheel. The force generated by this spring is determined by its stiffness and the maximum deflection of the track perpendicular to the tangent between the sprocket (or idler) and road wheel. The spring force is opposed by the track tension. Once the track tension is computed, the forces and moments transmitted to the sprocket (or idler) and road wheel can readily be determined.
b. The displacement of the road wheels with respect to each other will also cause track deflection. The resulting increase in track tension generates a restoring force on the displaced road wheel. This behavior is modeled by a vertically constrained, virtual linear bidirectional spring. If a road wheel is displaced relative to the adjacent one, the virtual spring will generate a vertical restoring force proportional to the relative displacement of adjacent road wheels.

An enhanced vehicle dynamics module of NRMM, VEHDYN 4.0, was released in 2008 [25].

It should be pointed out that existing ride dynamics analyses for tracked vehicles are all based on the assumption that the ground is rigid (nondeformable). Tests to validate these models have been conducted only on rigid surfaces with discrete obstacles or specific profiles. Figure 34.23 shows a comparison of the measured and predicted driver seat vertical acceleration of an armored personnel carrier with different track models traversing a 6′ obstacle on firm ground [22]. In Figure 34.23, track model #1 assumes the overall track tension being the initial track tension plus the tension due to the stretching of the track. In track model #2, the sag of the track segment between the sprocket and idler, represented by a quadratic polynomial, is incorporated in the evaluation of dynamic track load. Track model #3 is similar to track model #2, with the exception of the modeling of track sag, which is based on the classical catenary approach. Track model #4 is based on the model employed in the earlier version of VEHDYN mentioned previously.

It should be noted that while the ride quality of a tracked vehicle over rough, rigid surfaces may serve as a basis for comparing suspension designs, it does not represent the actual ride behavior of a tracked vehicle operating in the field. In an investigation carried out at Caterpillar into the vibration behavior of a D8L tracked tractor in the field, it has been found that the effects of dynamic stiffness and damping coefficient of the soil should be taken into consideration [26].

It should be mentioned that in recent years, a number of computer software packages for multibody dynamics, such as ADAMS and DADS, have become available. They have been used to

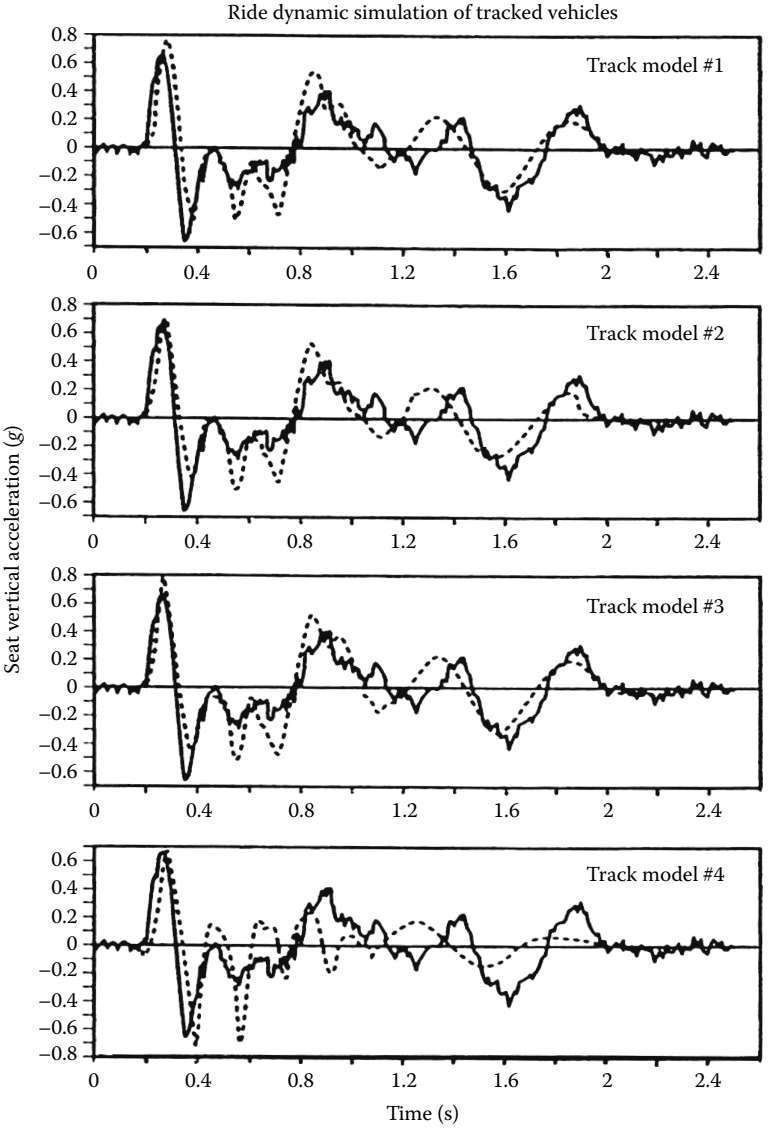


FIGURE 34.23 Comparison of the measured ride characteristics of an armored personnel carrier and those simulated using various types of track model over a 6 in. obstacle on a rigid surface (solid line for measured data and dotted line for simulated results). (Reproduced from Dhir, A. and Sankar, S., Analytical track models for ride dynamic simulation of tracked vehicles, *J. Terramech.*, 31(2), 107–138, 1994. With permission from the International Society for Terrain-Vehicle Systems.)

simulate the dynamic behavior of both wheeled and tracked vehicles over obstacles or on rough, rigid surfaces with encouraging results.

In summary, based on the brief discussion given earlier, it can be said that a great deal of effort has been expended on research into the ride dynamics of off-road vehicles over rough, rigid surfaces and considerable progress has been made. However, little research has so far been carried out on ride dynamics of off-road vehicles over deformable terrain, where the surface profile may be modified by vehicular load and the dynamic stiffness and damping of the terrain may have a significant influence on vehicle vibration.

34.5 CLOSING REMARKS

a. In recent years, comprehensive and realistic computer simulation models for predicting and evaluating the performance (mobility) of off-road wheeled and tracked vehicles over unprepared terrain have emerged. They take into account all major vehicle design parameters and pertinent terrain characteristics. The basic features of these models have been substantiated by field test data. It has been shown that these computer simulation models can provide more realistic and accurate predictions of vehicle performance (mobility) over unprepared terrain than empirical or semiempirical methods developed earlier.
b. A general theory for skid-steering of tracked vehicles on firm ground has recently been developed. It is shown that the turning behavior of tracked vehicles predicted using this new theory bears a strong resemblance to that measured in the field and that it provides significant improvements over existing methods. It is recommended that further research be conducted to extend this theory to the prediction and evaluation of turning behavior of tracked vehicles over deformable terrain where vehicle sinkage may be significant.
c. Significant progress has been made in recent years in the development of ride dynamics models for off-road vehicles, including both wheeled and tracked vehicles, over rough, rigid surfaces. Further studies should be carried out to extend these models to the prediction of ride dynamics of off-road vehicles over deformable terrain, where terrain profile may be significantly modified by vehicular load and the dynamic stiffness and damping of the terrain may have considerable effects on vehicle vibration.

Notations

B	tread (or track) of vehicle
b	width
c	cohesion
D	diameter
F	thrust
F_d	drawbar pull
F_i	thrust of inside track
F_o	thrust of outside track
f_r	coefficient of motion resistance
j	shear displacement
K	shear deformation modulus
l	contact length
M_r	moment of turning resistance
P	track pitch
p	pressure
R_c	motion resistance
r	radius
V	vehicle forward speed
V_j	slip velocity
V_t	vehicle theoretical speed
α	angle
δ_s	vehicle speed fluctuation
μ	coefficient of friction
μ_t	coefficient of lateral friction
τ	shear stress

ϕ angle of shearing resistance
ω angular speed
ω_i angular speed of inside track sprocket
ω_o angular speed of outside track sprocket

ACKNOWLEDGMENTS

Technical assistance provided by Dr. Wei Huang and Mr. Michael Galway in the preparation of this Chapter is greatly appreciated.

REFERENCES

1. Wong, J.Y., *Theory of Ground Vehicles*, 4th edn., John Wiley, New York, 2008.
2. Wong, J.Y., *Terramechanics and Off-Road Vehicle Engineering*, 2nd edn., Elsevier, Oxford, England, 2010.
3. Wong, J.Y., Garber. M., and Preston-Thomas, J., Theoretical prediction and experimental substantiation of the ground pressure distribution and tractive performance of tracked vehicles. *Proceedings of the Institution of Mechanical Engineers*, *Part D: Transport Engineering*, 198(D15), 265–285, 1984.
4. Wong, J.Y. and Preston-Thomas, J., Investigation into the effects of suspension characteristics and design parameters on the performance of tracked vehicles using an advanced computer simulation model. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Transport Engineering*, 202(D3), 143–161, 1988.
5. Wong, J.Y., Optimization of the tractive performance of articulated tracked vehicles using an advanced computer simulation model. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Journal of Automobile Engineering*, 206(D1), 29–45, 1992.
6. Wong, J.Y., Computer-aided methods for design evaluation of track systems. *Transactions of the Society of Automotive Engineers*, *Section 2, Journal of Commercial Vehicl*es, Paper No. 941675, 1994.
7. Wong, J.Y., Application of the computer simulation model NTVPM-86 to the development of a new version of the infantry fighting vehicle ASCOD. *Journal of Terramechanics*, 32(1), 53–61, 1995.
8. Wong, J.Y., Dynamics of tracked vehicles. *Vehicle System Dynamics*, 28(2 and 3), 197–219, 1997.
9. Wong, J.Y., Computer-aided methods for design evaluation of tracked vehicles and their applications to product development. *International Journal of Vehicle Design*, 22(1 and 2), 73–92, 1999.
10. Gao, Y. and Wong, J.Y., The development and validation of a computer-aided method for design evaluation of tracked vehicles with rigid links. *Proceedings Institution of Mechanical Engineers*, *Part D, Journal of Automobile Engineering*, 208(D3), 207–215, 1994.
11. Wong, J.Y. and Gao, Y., Applications of a computer-aided method to parametric study of tracked vehicles with rigid links. *Proceedings of the Institution of Mechanical Engineers, Part D, Journal of Automobile Engineering*, 208(4), 251–257, 1994.
12. Wong, J.Y., Optimization of design parameters of rigid-link track systems using an advanced computer-aided method. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Journal of Automobile Engineering*, 212(D3), 153–167, 1998.
13. Hayashi, I., Practical analysis of tracked vehicle steering depending on longitudinal track slippage. *Proceedings of the 5th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. II, Detroit, MI, 1975.
14. Steeds, W., Tracked vehicles. *Automobile Engineer*, 143–148, April 1950; 187–190, May 1950; and 219–222, June 1950.
15. Wong, J.Y. and Chiang, C.F., A general theory for skid-steering of tracked vehicles on firm ground. *Proceedings of the Institution of Mechanical Engineers*, *Part D, Journal of Automobile Engineering*, 215(D3), 343–355, 2001.
16. Ehlert, W., Hug, B., and Schmid, I.C., Field measurements and analytical models as a basis of test stand simulation of the turning resistance of tracked vehicles. *Journal of Terramechanics*, 29(1), 57–69, 1992.
17. Schwanghart, H., Measurements of forces on steered, towed wheels (in German). *Proceedings of the 7th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. I, Calgary, Alberta, Canada, 1981.
18. Gobbi, M., Aiolfi, M., Pennati, M., Previati, G., Levi, F., Ribaldone, M. and Mastinu, G., Measurement of the forces and moments acting on farm tractor pneumatic tyres. *Vehicle Systems Dynamics*, 43, Supplement (Tyre Models for Vehicle Dynamic Analysis), 412–433, 2005.
19. Dudzinski, P.A., Design characteristics of steering systems for mobile wheeled earthmoving equipment. *Journal of Terramechanics*, 26(1), 25–82, 1989.

20. International Standard Organization: ISO 2631/1-1985 (E), 1985; and ISO 2631-1: 1997 (E), 1997.
21. Lee, R.A. and Pradko, F., Analytical analysis of human vibration. *Transactions of the Society of Automotive Engineers*, 77, Paper No. 680091, 1969.
22. Dhir, A. and Sankar, S., Analytical track models for ride dynamic simulation of tracked vehicles. *Journal of Terramechanics*, 31(2), 107–138, 1994.
23. Bennett, M.D. and Penny, P.H.G., Dynamic simulation of track laying vehicles. *Proceedings of the 8th International Conference of the International Society for Terrain-Vehicle Systems*, Vol. I, 153–170, Cambridge, England, 1984.
24. Creighton, D.C., Revised vehicle dynamics module: User's guide for computer program VEHDYN II (Final Report), Technical Report SL-86-9, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, 1986.
25. Creighton, D.C., McKinley, G.B., Jones, R.A. and Ahlvin, R.B., Enhanced vehicle dynamics module: User's guide for computer program VEHDYN 4.0 (Final Report), U.S. Army Engineer Research and Development Center, Vicksburg, MS, 2008.
26. Huck, F.B., A case for improved soil models in tracked machine simulations. *Proceedings of the 1st North American Workshop on Modeling the Mechanics of Off-Road Mobility*, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS, 1994.

35 Motorcycle Handling Dynamics

Robin S. Sharp

CONTENTS

35.1 Introduction 1314
35.2 Rider Properties 1315
35.3 Theoretical Background 1316
35.4 Modeling the Motorcycle 1319
35.4.1 Minimum Requirements 1319
35.4.2 Computer-Aided Model Building 1319
35.4.3 Bodies and Freedoms for the Motorcycle 1320
35.4.4 Suspension and Steering Systems 1321
35.4.5 Tire/Road Contact Description 1321
35.4.6 Tire Forces and Moments 1322
35.4.7 Aerodynamic Forces and Moments 1324
35.4.8 Monoshock Suspension 1324
35.4.9 Chain Drive 1326
35.4.10 Speed and Steer Controllers 1327
35.4.11 Inputs 1328
35.4.12 Parameter Values 1329
35.4.13 Equilibrium State Checking and Power Balancing 1329
35.5 Motorcycle Behavior 1331
35.5.1 Equilibrium "Trim" States 1331
35.5.2 Stability for Small Perturbations from Straight-Running Trim States 1333
35.5.3 Normal Modes for Straight-Running Trim States 1338
35.5.4 Frequency Responses in the Neighborhood of Straight Running 1340
35.5.5 Stability in Cornering 1342
35.5.6 Cornering Modes 1345
35.5.7 Road Undulation Forcing 1345
35.5.8 Steering Compensation 1348
35.5.9 Acceleration and Deceleration 1349
35.5.10 Power Accounting 1354
35.6 Sensitivities to Configuration Perturbations 1355
35.6.1 Experimental Configuration Changes 1355
35.6.2 Theoretical Configuration Changes 1356
35.7 Optimal Linear Preview Rider Control and Handling Qualities 1359
35.7.1 Motorcycle Rider Model Background 1359
35.7.2 Existing Motorcycle Rider Models 1360
35.7.3 Outline of the Application of Optimal Linear Preview Regulator Theory 1361

35.7.4 Optimal Control Results 1364
35.7.5 Path Tracking Utilizing the Optimal Controls 1367
35.8 Conclusions 1372
Acknowledgments 1373
References 1373

35.1 INTRODUCTION

Motorcycle handling involves both machine and rider. It is a man–machine problem. The rider behavior probably can be influenced by education and training but only a little, so the main engineering focus is on the machine. However, the machine properties need to be aligned closely with the needs and preferences of the rider. Consequently, the relevant rider characteristics are reviewed first. Then, the behavioral features of motorcycles are discussed.

The view is taken that a high level of agreement between measured and computer-simulated responses of motorcycles and riders to steering control and other inputs has already been established [1–22], and it is therefore possible to use advanced computer model predictions to explain and to understand the real thing. This is not to imply that predicting the responses of a particular machine–rider combination to path demands or other control inputs over a wide spectrum of operating conditions with accuracy is a trivial matter. Such a task needs a very large expenditure on model building and parameter measurement, much of the latter demanding special equipment, as a minimum. However, in general terms, advanced model predictions can be regarded as sufficiently accurate to properly reveal much of how motorcycles respond to control inputs.

Ideal responses cannot be specified, although minimizing the lag between yaw rate response and steer torque input in single- or double-lane-change maneuvers, as revealed by the peak of the cross-correlation coefficient, has been shown to correlate with subjective good handling [14]. Other desirable properties have been suggested [19,21], while, for racing, minimum-time capability is clearly ideal. Indeed, for general use, it must be the case that sports machines, cruisers, and commuter bikes represent different ideals. However, all reasonably conventional motorcycles have some stability and response properties in common, which have to be understood and controlled. At very low speeds, an uncontrolled motorcycle will diverge in roll in particular and will fall over or capsize. At medium or high speeds, depending on the design, the steering system is likely to wobble, while at high speeds, a rolling/yawing/steering oscillation, weave, is likely to show.

In Section 35.2, rider properties are discussed and some deductions about motorcycles are made. In Section 35.3, a general theoretical background is provided. The background is used to make a connection between steady-turning behavior and static stability, and a comparison with the better-known car situation is made. In Section 35.4, there is a discussion of the fundamental needs of a mathematical model that is capable of representing real behavior in some detail. That background is used to explain the building of such a model. Some procedures for checking the health of the model are described. Section 35.5 contains behavioral results from the model. Steady-turning equilibrium "trim" states are shown, and eigenvalues and eigenvectors generated for small perturbations from trim are calculated. Frequency responses for straight running are given and stability in cornering is considered. Responses to regular road undulation forcing are computed and related to loss-of-control accidents. The possibility of replacing the standard steering damper with a more general mechanical or electromechanical steering compensator is mentioned, and the influences on stability of acceleration and deceleration are studied. The chapter concludes with some power accounting.

Section 35.6 is concerned with sensitivities to design parameter changes. Firstly, practical results from the literature are discussed. Then the motorcycle model is employed to predict behavioral changes resulting from small variations in a number of parameters. Section 35.7 is about rider

control, concentrating on optimal linear quadratic preview control, which is a recent development. This is believed to provide insight into how riders use knowledge of the machine dynamics and of the path to be followed to exercise effective steering control, especially near to straight-running conditions. Conclusions are stated in Section 35.8.

35.2 RIDER PROPERTIES

The rider is a significant part of the structure of a motorcycle, in addition to being the controller. Lateral control can be exercised by steering, body leaning relative to the motorcycle frame, and lateral body movement. Rider relative displacements may be consequences of motorcycle motions or of rider muscular control activity. Such control activity is often passive, insofar as it occurs pre-maneuver, and is apparently designed to obtain some advantage in the later maneuver. The rider's role as a part of the structure is most important when the machine is lightest and the rider is heaviest.

The responses of several riders as passive structural components to motions of a mock motorcycle frame without longitudinal movement were measured in the laboratory by Nishimi et al. [13]. A two-degree-of-freedom linear model of the rider was found to fit the results well. The freedoms involved lateral translation at the hips and rolling of the upper body relative to the frame, to which sinusoidal rolling motions were applied. Typical natural frequencies and damping factors of a rider model fitting the results were 4.2 Hz and 0.40 for the hip translation mode and 1.1 Hz and 0.47 for the body lean mode. Fortunately, since riders vary over a wide range, these values do not critically influence the behavior of motorcycle/rider combinations. They appear to provide a good basis for modeling a generic rider as a structure. Riders also have some capability to use muscle tensing to change their structural properties [13,24,25]. Experimental results to date suggest that such muscle tensing allows the rider to increase stiffnesses by at least an order of magnitude but to increase effective damping coefficients only a little [24]. The mobility of the rider's upper body provides some useful damping of the oscillatory modes of the system*, suggesting that the lightest riders represent worst-case configurations from this point of view.

From a controller viewpoint, the rider can use both body and arm muscles to cause body or steering inputs within the control bandwidth, with steer torque inputs being reacted on the rider's upper body. This is genuine active control. The bandwidth will increase with skill level, but it is likely that, even under racing conditions, very little control power is used above 0.5 Hz [26]. The rider can also use near-zero frequency control to effectively alter the higher frequency dynamics of the motorcycle/rider system. When the rider substantially alters his or her body position relative to the motorcycle for drag reduction or vigorous cornering, this is regarded as parameter adaptation, not as active control. The equilibrium or trim states of the system are affected and the dynamics involving small perturbations from trim are correspondingly influenced. When the rider finds favorable influences, reinforcement learning [27] will contribute to making the changes into a habit.

The rider may operate in "fixed control," effectively using steer angle and/or upper body lean angle as controlled variables, or alternatively, he or she may use "free control" in which steer torque and lean torque are the controlled variables. In principle, the same options exist in relation to longitudinal control by brake and throttle actions, but conventional arrangements more or less compel throttle position control and brake force control to be utilized by riders [28].

It is evident that riders are learning and optimizing controllers. They learn by repeated trials and observations, and they optimize some combination of system performance like path following accuracy and speed maintenance and control effort. In a learning phase, fixed- and free-control strategies are no doubt evaluated, to see which is easier and more effective. In application, the control mode and the concentration level chosen will vary with the circumstances from high effort and high accuracy at the one extreme to the converse. In the trials, the relative effort and

* To be elaborated in Section 35.5.2.

reward associated with body torque and steer torque controls will be assessed, it being apparent, at the outset, that using body muscles for active control is likely to be feasible only at very low frequencies. For most people, a realistic estimate of the bandwidth for body control is perhaps 0.1 Hz, rather than the 0.5 Hz associated with steer torque control. It is surmised, at this stage, that riders use feedback control for stabilization and disturbance rejection and preview control for path following and speed maintenance. Provided that the available steering torque and/or rider body lean torque actuations are effective, as will normally be the case, stabilization of divergent motions is possible if the time constant is sufficiently long and stabilization of oscillatory modes by active control is possible if the frequency of the oscillations is low enough in relation to the control bandwidth.

Sometimes, it is argued that good stability margins lead to poor maneuverability [29] and therefore that stability can be excessive. However, maneuverability depends fundamentally on the relationship between the external forces that can be called upon by the rider and the responsiveness of the motorcycle to those forces. This is substantially independent of the stability properties. The rider has more or less direct control of the lateral force at the front tire contact with the ground. From a straight and level condition, maximum response is obtainable by steering the front wheel quickly to the condition at which it will give maximum side force. Most riders are not brave enough or skillful enough to do this. Typically, they will extract much less than the full performance capability of their machine through conservative riding [30]. Then, the extent to which the machine will respond to the external forces imposed will depend on its mass and inertia properties, its wheelbase, and many design details, combining in a complex way that can be understood only by examining the full nonlinear system dynamics at limiting performance. It seems likely that good stability margins will actually make a positive contribution to maneuverability, by encouraging riders to fully use the performance envelope, without fear of losing control. Stability is an important theme in the discussion to follow.

35.3 THEORETICAL BACKGROUND

The equations of motion of a mechanical system are representable in the form [31]

$$\left[S(\underline{q},t)\right]\dot{\underline{q}} = \underline{v}(\underline{q},\underline{u},t) \tag{35.1}$$

for the kinematics and

$$\left[M(\underline{q},t)\right]\dot{\underline{u}} = \underline{f}(\underline{q},\underline{u},t) \tag{35.2}$$

for the dynamics. The equations of vehicle dynamics can normally be reduced to the form

$$\dot{\underline{x}} = \underline{f}(\underline{x}) + \underline{g}(\underline{u}) \tag{35.3}$$

in which $\underline{x}$ is the state vector, $\underline{u}$ is the input vector, and $\underline{f}$ and $\underline{g}$ are vector functions. For a motorcycle, the input vector $\underline{u}$ can be thought to include steering torque (or steer angle), rider lean torque, and throttle position or brake force. Front and rear brakes can normally be operated independently, but we will consider that to be too detailed a level for the present discussion. Also, later, consideration will be given to the rider's choice of fixed or free control and the relative importance of steer torque and lean torque. The functions involved in the equations will normally be continuous and differentiable. The system is likely to be smooth. Without losing any of the system dynamics, the states can be chosen in such a way that they all have zero values in a steady-state trim condition. Thus, we omit from consideration the longitudinal and lateral positions, the attitude angle, and the wheel spin and drive sprocket rotation angles.

A straightforward use of the equations of motion is to numerically integrate them for given initial conditions and defined control inputs, but such usage is not so cost-effective in terms of understanding the dynamics and the machine design. Linearization is highly desirable. Linear system theory is extremely effective in allowing behavior to be efficiently characterized and design and behavior to be related to each other.

The common linearization involves series expansion about a trim or equilibrium state, although other forms exist [32]. Trim states are solutions of

$$\underline{f}(\underline{x}) = -\underline{g}(\underline{u}) \tag{35.4}$$

solutions that imply constant speed running. Supposing the system order to be n and there to be m inputs, the constant-coefficient small-perturbation equations are of the form

$$\dot{\underline{x}} = \left[A\right]\underline{x} + \left[B\right]\underline{u} \tag{35.5}$$

in which the elements of $\left[A\right]$, the (n^*n) Jacobian matrix of derivatives $\underline{\partial f}/\underline{\partial x}$ and $\left[B\right]$, and the (n^*m) matrix $\underline{\partial g}/\underline{\partial u}$ depend on the trim state. Trim-state solutions may be plotted in a handling diagram form, conventionally consisting of lateral acceleration response to steer torque or steer angle input for a fixed speed and zero body lean torque [33]. The small-perturbation-from-trim equations are linear with constant coefficients. The lateral equations contain the forward speed and the rotor-spin velocities for the trim state as parameters. If and when the gradient of a handling curve such as that described becomes infinite, it is implied that the system at the trim state involved has a zero eigenvalue [28]. If the handling curve employs steer torque as a reference, it is the free-control motorcycle that has the zero root, while if steer angle is the reference, it is the fixed-control machine that has the zero root. These implications can be demonstrated by the following argument.

Suppose that $\underline{x}_0$ is a trim state including lateral acceleration α_0 and steer angle δ_0, corresponding to steer torque input τ_0 and speed v_0, and that a neighboring trim state is characterized by $\underline{x}_1$, α_1, δ_1, τ_1 and the same v_0. Due to the closeness of the second trim state to the first one, the second can be obtained from the first by way of the linearization of the system about the first trim state.

Suppose that $\left[A\right] = \left[A_0\right]$ and $\left[B\right] = \left[B_0\right]$ for the first trim state. Then the change in steady-state response due to a change in steer torque input is given by

$$(x_1 - x_0) = -\left[A_0\right]^{-1}\left[B_0\right](\tau_1 - \tau_0) \tag{35.6}$$

If the running conditions for the trim state are such that $\left[A_0\right]$ is singular, that is, if $\det\left[A_0\right] = 0$, the increment $(x_1 - x_0)$ is infinite. In particular, the gradient of the lateral acceleration versus steer torque handling diagram becomes infinite for such circumstances. Correspondingly, if the control input were steer angle instead of steer torque, the matrices $\left[A_0\right]$ and $\left[B_0\right]$ would belong to the fixed-control motorcycle.

The stability of the first trim state is indicated by the eigenvalues of $\left[A_0\right]$ and the roots λ of the characteristic equation, a polynomial in λ:

$$\det\left(\left[A_0\right] - \lambda\left[I\right]\right) = 0 \tag{35.7}$$

The zero-order term in the expansion of the determinant is $\det\left[A_0\right]$, and if this term has value zero, $\lambda = 0$ is a solution. Thus, a zero eigenvalue arising from the singularity of the characteristic matrix at a particular operating condition must be accompanied by an infinite slope on the handling diagram for the matching problem at the operating point.

If there exists some problem-specific knowledge, it may be possible to associate the infinite slope on the handling diagram with a transition from statically stable to statically unstable running of the system as it passes through the operating point. This is the case with the well-known yaw/sideslip car model [33], in which the low order of the system and the inherent stability of the car at low speeds make it certain that the infinite slope on the handling diagram does indeed indicate a transition to static instability. However, in the absence of such problem-specific knowledge and especially when the system of interest is of high order, the relationship between the steady-state responses and the static stability is less certain.

For any running condition making $\left[A_0\right]$ singular, $\det\left[A_0\right]$ will be zero and the system will have a zero eigenvalue. The characteristic equation of the system can then be written

$$\lambda.G(\lambda) = 0 \tag{35.8}$$

The roots of $G(\lambda) = 0$ may include positive reals, indicating monotonic divergence, or complex conjugate pairs with positive real parts, indicating oscillatory instability. Thus, there may be little significance in the zero eigenvalue in a relatively complex case. Great care needs to be exercised in translating ideas from one vehicle sector, like understeer and oversteer, for example, and applying them directly to another one. The steady-turning responses of a motorcycle are much less indicative of the stability than for a car.

In general, the longitudinal and lateral equations are coupled but, if the trim state involves straight running and the machine has right/left symmetry, the equations decouple into an in-plane set and an out-of-plane set [22,33,34]. The eigenvalues and eigenvectors of $\left[A_0\right]$ indicate the stability, natural frequency, and mode-shape properties of the small-perturbation system, and the frequency responses to inputs $\underline{u}$ show the response properties in the neighborhood of the relevant trim state [35].

If responses $\underline{y} = \left[C\right]\underline{x}$ are considered, the corresponding frequency response functions are given by

$$\underline{H}_0 = \left[C\right]\left(\lambda\left[I\right] - \left[A_0\right]\right)^{-1}\left[B_0\right] \tag{35.9}$$

and in the case of road profile excitation, when the rear input can be considered a time-delayed version of that at the front, responses can be combined using

$$h(j\omega) = h_f(j\omega) + e^{-j\omega\tau}h_r(j\omega) \tag{35.10}$$

in which τ is the wheelbase travel time.

For variable speed operation, linearization about a trajectory is possible [32]. Such linearization will yield a time-varying system, the stability and response properties of which are much more difficult to determine. However, if the rate of speed variation is sufficiently small, frozen-time theory yields a close approximation to exact behavior [32,36] and the difficulties associated with time-varying systems can be avoided. Constant-coefficient linear system analysis continues to be useful. The influences of acceleration or deceleration can be calculated with reasonable accuracy through frozen-time theory.

The linearizations described take into account only the system dynamics in the neighborhood of trim states. Away from these neighborhoods, there is no guarantee of accuracy. However, it is reasonable to take the view that, with a well-behaved man–machine system, the control by the man will be of sufficient quality, in the great majority of operating circumstances, that the motions will remain close to trim and that therefore the dynamics that matter will be computed properly by the methods described. If a rider wants to explore running conditions far from equilibrium, to develop skill in dealing with them, for example, large-amplitude transient control inputs should be applied with well-chosen phasing of the input stages to yield reinforcement of the successive responses. Similarly, in the theoretical world, large-amplitude transient control inputs are required to excite extreme responses, far away from trim states.

35.4 MODELING THE MOTORCYCLE

35.4.1 Minimum Requirements

Although quite simple models may be capable of representing low-speed and low-frequency behavior well, generally effective mathematical models of motorcycles are necessarily not simple. Essential freedoms to be included at the outset are lateral translation, yaw, roll, steer*, and frame twisting [22,37,38]. Some machines may have significant torsional compliance in the rear wheel attachment, which is known to destabilize the weave mode [39], but modern machines are usually stiff enough in this regard (stiffness >12 kN m/rad) to be treated as rigid. The rider also should be treated as compliant, although this issue is not so critical, with the characterization of [13] being the most compelling available. Tire lateral forces, aligning moments, and overturning moments in response to tire load change, longitudinal slip, sideslip, and wheel camber should be included. Also, tire forces and moments must be considered dynamic, rather than static [22,40]. Time lags in the tire force and moment system occur primarily because the tire sidewall structure is compliant. If these time lags are omitted from consideration, oscillatory modes are predicted wrongly [22,40]. It is essential to use an automated model builder for a system of this complexity [41].

35.4.2 Computer-Aided Model Building

AutoSim [42] is just one of a number of multibody modeling systems currently in common use. AutoSim is a set of LISP macros, enabling the description of mechanical multibody systems, with the addition of nonmechanical subsystems, if necessary. The syntactic rules of AutoSim are straightforward. The output from AutoSim takes one of several forms: (a) a Rich Text Format file containing the symbolic equations of motion of the system described, (b) a "C" or "FORTRAN" language simulation program with appropriate data files containing parameter values and simulation run control parameters (or Simulink S-function), or (c) linear state-space equations in a MATLAB® "M-file" format. AutoSim's linearization of the nonlinear equations of motion is symbolic and completely general. The linearization involves series expansions of smooth functions about trim states. From the general equations of motion, it is possible to determine trim states and all the parameters required to describe fully a small-perturbation dynamics problem. The multibody basis of AutoSim is Kane's equations [42], an expression of the virtual power or Jourdain principle [43], which deals equally easily with holonomic and non holonomic constraints. The simulation programs written by AutoSim contain many software engineering features that help to reduce execution times.

At the core of the language are commands like add body, add point, add speed constraint, add line force, add strut, add moment, and no movement, which, to some extent, are self-explanatory. Such commands contain argument lists that are required to make their meanings precise. A program begins by defining an inertial reference frame `n,` with fixed origin, `n0,` and fixed directions `[nx], [ny],` and `[nz]`. As new bodies are added, having freedoms relative to n, local origins and axes are employed. All the reference axis systems introduced are right-handed Cartesian sets, and inertias are normally referenced to axes parallel to these through the material body's mass center. The code can transform from local coordinates to global ones and vice versa. Points specified globally are conveniently used to define points in bodies. The body-fixed points coincide with the corresponding global points, fixed in `n,` when the system is in its nominal configuration. Most points are fixed in bodies, but a point may be defined as "moving" with its location in a body determined from its specified coordinates. This is especially useful for describing tire/ground contact.

Position constraints, velocity constraints, or both together can be added. The analyst has the option to specify which freedom should be eliminated from the problem as a result of each constraint added. Constraints may require points on two bodies to have no relative movement, for

* To analyze the fixed-control properties of the motorcycle, the steer degree of freedom is omitted [22,40].

example, or they may consist of symbolic expressions that must have zero value. To deal with a closed kinematic loop, this is first opened by an imagined cut in such a way that the longest chain of bodies contributing to the loop is as short as possible. The loop closure constraints are then added.

There are several commands for defining and manipulating vectors. For example, `pos(p1,p2)` defines a vector that points from `p2` to `p1`, `dot(vel(p1),[nz])` is the component of the absolute velocity of `p1` in the direction `[nz]`, `dir(pos(p1,p2))` is a unit vector in the direction `p2` to `p1`, `dplane([rwy],[nz])` is the projection of `[rwy]` onto the plane normal to `[nz]`, and `cross(a,b)` is the vector product of a and b. In each case, the square brackets indicate a unit vector along a (Cartesian) body axis. The "setsym" command is used to define a new symbol, such a symbol being identified subsequently by the "@" character.

New state variables can be added to a model and they can be made integral functions of existing variables. Differentials with respect to time and partial differentials of variables can be defined. Max, min, sign, and ifthen functions can be used, allowing discontinuous actions, like tires leaving the ground and limit stops being contacted, to be modeled easily. If necessary, the analyst can define how discontinuous functions are to be differentiated. Small-angle theory can be employed to reduce simulation times where it is considered appropriate.

At the head of an AutoSim program are some commands that are used to reset the system, define the gravitational field, select the unit system to be employed, specify the simulation language in which the code written by AutoSim is to be generated, and select the type of numerical integrator to be installed in the code and to control the symbolic manipulations in various other ways. When the analyst does not exercise choice overtly in these matters, default settings come from a configuration file that the analyst is free to access.

35.4.3 Bodies and Freedoms for the Motorcycle

Figure 35.1 shows a modern high-performance motorcycle diagrammatically. Separate masses are associated with the rear wheel, the gearbox sprocket, the swing arm, the rider's lower body, the rider's upper body, the main frame, the front frame, and the front wheel. The main frame has the freedom to translate in any direction and to yaw, pitch, and roll, while the front frame has relative twist and steer freedoms. The rider has lower body lateral translation and upper body relative roll freedoms. The rear suspension is by swing arm and monoshock spring/damper system, and the rear wheel is driven by chain and sprocket. The gearbox sprocket is given a spin inertia that accounts for its connection to the rotating masses in the gearbox and engine. The bodies and their relationships to each other are shown in Figure 35.2. Initially, the bodies, joints, and freedoms implied by Figures 35.1 and 35.2 and all the key points in the system's nominal state are defined.

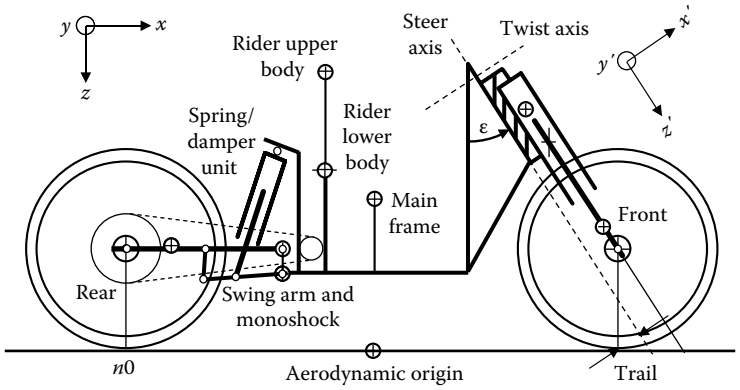


FIGURE 35.1 Diagrammatic motorcycle showing some geometric features.

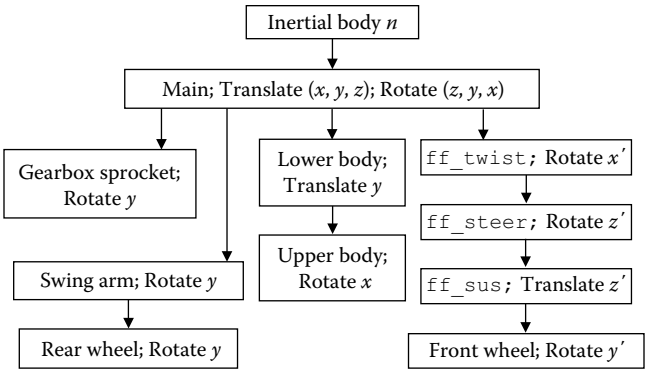


FIGURE 35.2 Body parent/child structure and relative degrees of freedom for motorcycle model. Axis directions are indicated in Figure 35.1.

35.4.4 Suspension and Steering Systems

Spring/damper units for suspensions, frame compliance, rider lower and upper bodies, and a steering damper are added. The monoshock rear arrangement requires special consideration and it is explained further in Section 35.4.8. Limit stops are associated with each suspension and the steering system. Forces or moments arising from limit stops become nonzero only when deflections exceed defined values and they are usually made a high power of the deformation of the stop, so that the force or moment rises from zero steeply once the stop is contacted.

35.4.5 Tire/Road Contact Description

The geometry of the contact between the front tire and the ground is a relatively complex part of the motorcycle modeling. It is also important to the behavior of the machine. To define each tire/ground contact point (Figures 35.3 and 35.4), the vertical and the wheel spindle directions are used in a vector (cross) product to describe the longitudinal direction, with respect to the wheel. Similarly, the wheel radial direction, OC in Figure 35.3, comes from combining the longitudinal and wheel spindle directions. The vector OC is of fixed length and so is completely specified. *G* is vertically

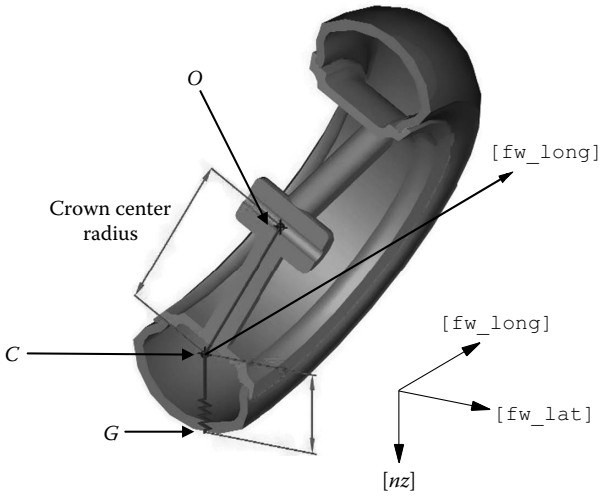


FIGURE 35.3 Diagrammatic 3D front wheel contact geometry showing the possible movement of the contact center circumferentially and laterally. The contact point is vertically below the point *C*.

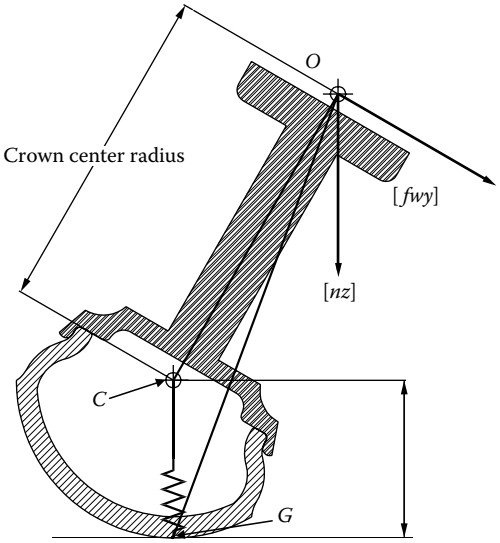


FIGURE 35.4 2D tire-to-ground contact geometry, showing the lateral movement of the contact center and the vertical flexibility of the tire structure. The change from nominal in the compression of the tire spring and the radial stiffness, together with the nominal load, allows the wheel load to be tracked.

below C, and the difference between the tire crown radius and the distance $\underline{CG}$ defines the change in the tire carcass compression from the nominal state and hence the change of the wheel load from the nominal, via the tire radial stiffness. If the road is profiled, the road height is accounted for in working out the wheel load. The vector $\underline{OG} = \underline{OC} + \underline{CG}$ defines the contact point, which belongs to the wheel but moves within it. G remains at road surface height, but the tire load cannot become negative. If the tire leaves the ground, the shear forces are zero, whatever the other conditions are. Tire forces are applied to the point G, in each case.

35.4.6 Tire Forces and Moments

The basis for the tire force and moment descriptions here is the "Magic Formula" [44–46]. The original development was for car tires [33], in which context, it has become dominant. The extension to cover motorcycle tires involves substantial changes to accommodate the completely different roles of sideslip and camber forces in the two cases. The "Magic Formula" is a largely empirical set of equations relating load, slip ratio (longitudinal slip), lateral slip, and camber angle to longitudinal force, side force, and aligning moment (and possibly overturning moment), with constraints on the parameters to prevent the behavior from becoming unreasonable in any operating conditions. Only very limited parameter values can be found in the literature, but a certain amount of relevant experimental data are available. Such data have been used for full Magic Formula parameter identifications for contemporary high-performance tires [47]. The parameter values for a given tire and the relevant equations [33,47] allow the calculation of the steady-state force and moment system for any realistic operating condition. Examples of experimental and fitted results are shown in Figures 35.5 through 35.7. Many more details are given in [47].

The tire force and moment model needs to be supplied with wheel load, longitudinal slip, lateral slip, and wheel camber angle. The wheel load comes from the nominal load, the tire radial stiffness, and the tire compression relative to the nominal state. The longitudinal slip is the rearward component of the material contact point velocity divided by the absolute value of the rolling velocity, the latter being the forward velocity of the contact point (or the crown center point, since these are the same). The contact point is defined by its coordinates in the parent body of the wheel, and it is de spun relative to the material contact point. Thus, the longitudinal slip is given by an expression of the form.

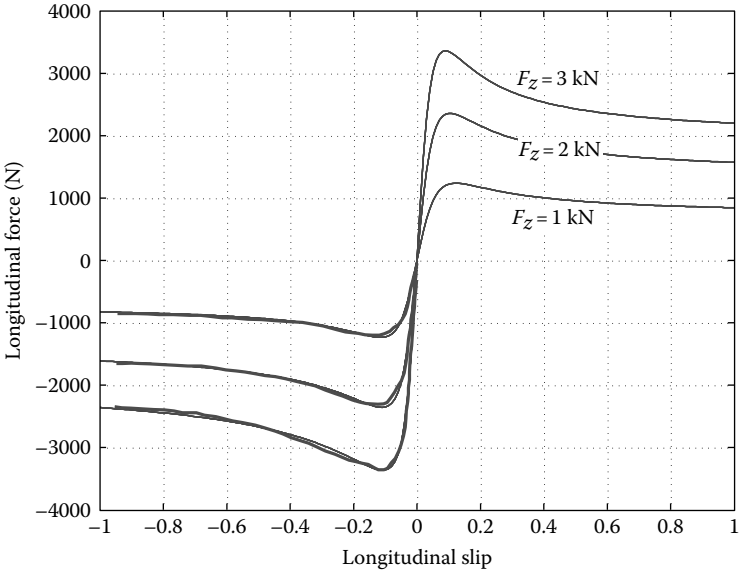


FIGURE 35.5 Tire longitudinal force results for a 160/70 tire (From Pacejka, H.B., *Tyre and Vehicle Dynamics*, Butterworth Heinemann, Oxford, U.K., 2002) (thick lines) with best-fit reconstructions by Magic Formulae (thin lines), for loads of 1, 2, and 3 kN.

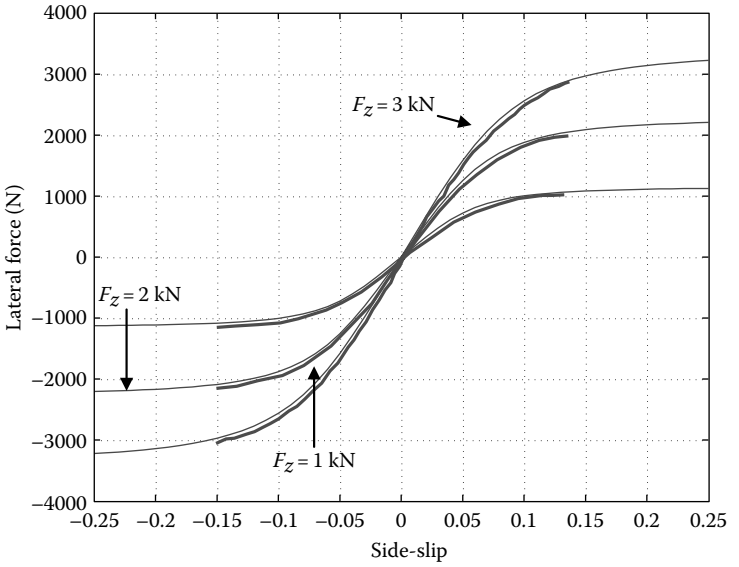


FIGURE 35.6 Tire lateral force results for a 160/70 tire at 1, 2, and 3 kN loads (From Sharp, R.S. etal., *Multibody Syst. Dyn.*, 12(3), 251, 2004 (thick lines) with best-fit reconstructions by Magic Formulae (thin lines).

$$\kappa = \frac{-(\text{rolling velocity} + \text{spin component of longitudinal velocity})}{\text{abs(rolling velocity)}}$$

The lateral slip is the ratio of the (negative) lateral velocity of the tire contact center point to the absolute value of the rolling velocity, and the wheel camber angle is the arcsine of the vertical component of the wheel spindle unit vector. A complete set of parameter values for a given, presumed symmetric tire [47] allows the calculation of the steady-state force and moment system for any realistic operating condition.

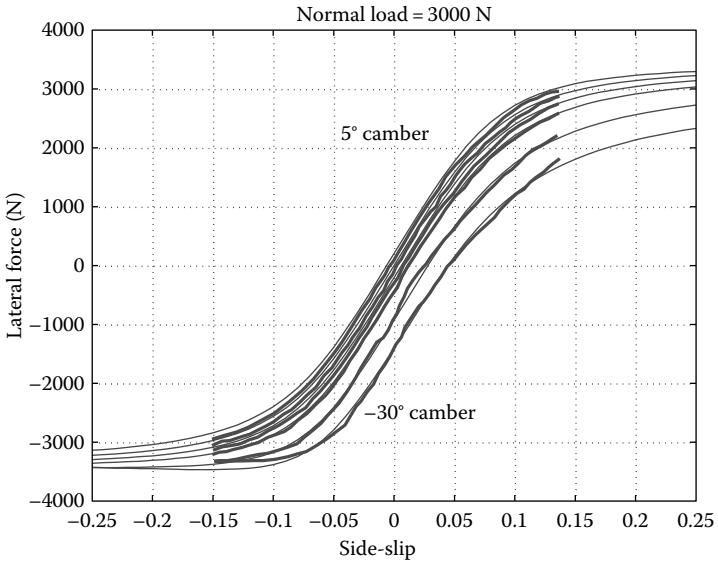


FIGURE 35.7 Tire lateral force results for a 160/70 tire at 3 kN load and camber angles of 5°, 0°,–5°,–10°,–20°, and–30° (From Sharp, R.S. etal., *Multibody Syst. Dyn*., 12(3), 251, 2004) (thick lines) with best-fit reconstructions by Magic Formulae (thin lines).

35.4.7 Aerodynamic Forces and Moments

Aerodynamic drag, lift, and pitching moment data come from a Triumph motorcycle of similar style and dimensions to the GSX-R1000 [47]. This is steady-state drag force, lift force, and pitching moment data from full-scale wind tunnel testing, with results for a prone rider being adopted. If the rider were upright, the frontal area would be larger by 20% and the drag coefficient would be larger by 15%. A fuller account of aerodynamic force and moment coefficients can be found in [14].

35.4.8 Monoshock Suspension

The motorcycle rear suspension arrangement is shown diagrammatically in Figure 35.8. It uses a single spring/damper unit with a mechanical linkage connection to the swinging arm. Many modern rear suspensions are of this type, although several variants of it exist [48]. It involves a closed kinematic loop. Such a suspension can be modeled literally, link by link and joint by joint, or via a separate geometric pre-analysis. Such a pre-analysis yields an analytic relationship between the swing-arm angle change and the moment of the spring force about the swing-arm pivot, which is used directly in the multibody model building. Alternatively, if the pre-analysis were too complex to give an analytic result, a numerical relationship between the angle and the moment could be found. This could be replaced by an approximate functional

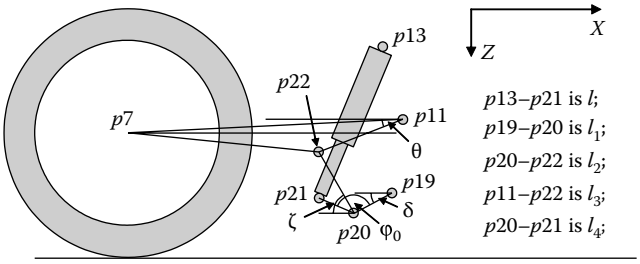


FIGURE 35.8 Geometry of monoshock suspension arrangement on GSX-R1000 motorcycle. Distances between various points are also defined in the diagram.

relationship covering the practical range of the swing-arm movement. A low-order polynomial will usually suffice [48]. Literal modeling is the simplest, but it may provide equations of motion that integrate relatively slowly, since the simulation has to solve the kinematic loop equations at each integration step [48].

An analysis from [48] follows.

Points $p11$, $p13$, and $p19$ are fixed to the main frame. l_1, l_4, and φ_0 are dimensions of the swinging link and l_2 the length of the pull rod. The length l_3 in the swing arm is fixed. The spring/damper unit is of variable length l. θ is the angle of the swing arm to the horizontal x-axis, while δ is the corresponding angle for the swinging link. Traversing the loop $p11$-$p22$-$p20$-$p19$-$p11$, both x and z displacements are nil, since we end where we begin. Therefore,

$$x_{11} - x_{19} - l_3 \cos\theta + l_2 \cos\zeta + l_1 \cos\delta = 0 \tag{35.11}$$

$$z_{11} - z_{19} + l_3 \sin\theta + l_2 \sin\zeta - l_1 \sin\delta = 0 \tag{35.12}$$

Forming l_2^2 as $(l_2^2 \sin^2\zeta + l_2^2 \cos^2\zeta)$ and substituting

$$c_1 = -x_{11} + x_{19} + l_3 \cos\theta \tag{35.13}$$

$$c_2 = -z_{11} + z_{19} - l_3 \sin\theta \tag{35.14}$$

we obtain

$$l_2^2 = (c_1 - l_1 \cos\delta)^2 + (c_2 + l_1 \sin\delta)^2 \tag{35.15}$$

from which it can be shown that

$$\delta = \arcsin\left(\frac{l_2^2 - l_1^2 - c_1^2 - c_2^2}{2l_1\sqrt{c_1^2 + c_2^2}}\right) + \arctan\left(\frac{c_1}{c_2}\right) \tag{35.16}$$

which is a function of θ only. Also,

$$x_{21} = x_{19} - l_1 \cos\delta + l_4 \cos(\phi_0 + \delta) \tag{35.17}$$

$$z_{21} = z_{19} + l_1 \sin\delta - l_4 \sin(\phi_0 + \delta) \tag{35.18}$$

with

$$l = \sqrt{(x_{13} - x_{21})^2 + (z_{13} - z_{21})^2} \tag{35.19}$$

so that l can be found as a function of θ, $l = f_1(\theta)$ say, by substitution for x_{21} and z_{21} in this expression. Figure 35.9 illustrates the outcome. If a small change $\delta\theta$ in θ occurs, in which the corresponding change in l is δl, the moment M corresponding to a spring/damper force $f_2(l, \dot{l})$ is $f_2(l, \dot{l}).dl/d\theta$ by virtual work. The properties of the spring/damper unit can thus be expressed in terms of an equivalent moment $M(\theta, \dot{\theta})$ about the swing-arm pivot, as

$$M = f_2\left\{f_1(\theta), \frac{df_1(\theta)}{d\theta}\dot{\theta}\right\}\frac{df_1(\theta)}{d\theta} \tag{35.20}$$

which can be fully automated.

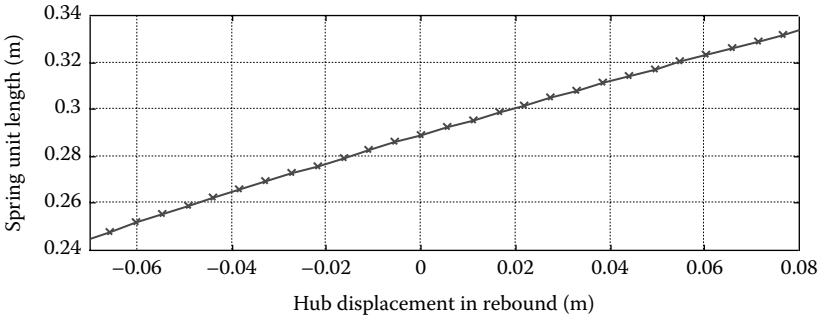


FIGURE 35.9 Spring/damper unit length to wheel displacement relationship for GSX-R1000 motorcycle.

35.4.9 Chain Drive

The chain functions as a tension link between the points p_{gu} on the gearbox sprocket and p_{wu} on the rear wheel sprocket when the engine is driving the wheel and between p_{wl} and p_{gl} when the wheel is driving the engine (Figure 35.10). Some slack or clearance is allowed between driving and overrunning, and the chain is treated as elastic and damped. The force transmitted between the relevant points thus depends on some logic, to decide between drive and overrun, and the relative angular displacements and velocities of the gearbox sprocket and the rear wheel. The points themselves need to be defined as moving points in the appropriate rotor. The point definitions are derived via a geometric analysis from [48] that follows.

From Figure 35.10, the following relationships can be deduced:

$$\mathbf{r}_{sa} = -r_{sa}\cos\theta\,\mathbf{i} + r_{sa}\sin\theta\,\mathbf{k} \tag{35.21}$$

$$\mathbf{r}_{s0} = (x_{23} - x_{11})\,\mathbf{i} + (z_{23} - z_{11})\,\mathbf{k} \tag{35.22}$$

$$\mathbf{r}_{wu} = -r_w \sin\eta_u\,\mathbf{i} - r_w\cos\eta_u\,\mathbf{k} \tag{35.23}$$

$$\mathbf{r}_{gu} = -r_g \sin\eta_u\,\mathbf{i} - r_g\cos\eta_u\,\mathbf{k} \tag{35.24}$$

$$\mathbf{r}_{wl} = r_w \sin\eta_l\,\mathbf{i} + r_w\cos\eta_l\,\mathbf{k} \tag{35.25}$$

$$\mathbf{r}_{gl} = r_g \sin\eta_l\,\mathbf{i} + r_g\cos\eta_l\,\mathbf{k} \tag{35.26}$$

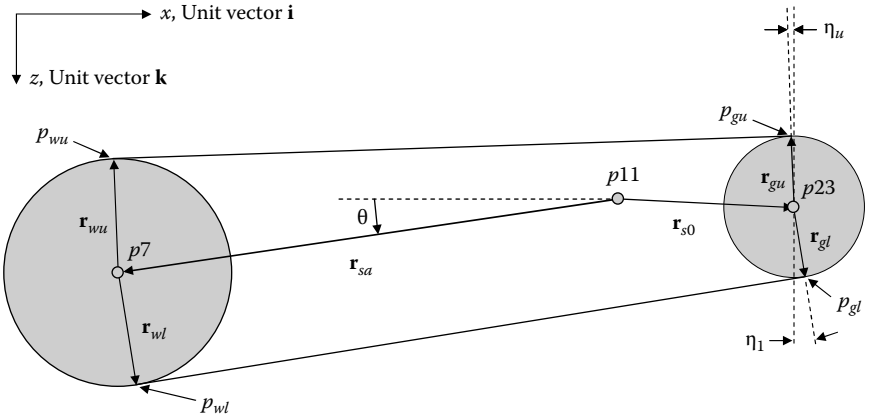


FIGURE 35.10 Geometry of chain drive with swinging-arm suspension. The point $p11$ is the swing-arm pivot center and $p23$ is the gearbox sprocket center.

The chain lines are perpendicular to the radius vectors, so that

$$(-\mathbf{r}_{wu} - \mathbf{r}_{sa} + \mathbf{r}_{s0} + \mathbf{r}_{gu}).\mathbf{r}_{gu} = 0 \tag{35.27}$$

$$(-\mathbf{r}_{wl} - \mathbf{r}_{sa} + \mathbf{r}_{s0} + \mathbf{r}_{gl}).\mathbf{r}_{gl} = 0 \tag{35.28}$$

Through some algebra, these equations yield

$$\eta_u = \sin^{-1}\left(\frac{r_g - r_w}{\sqrt{(r_{sa}\cos\theta + x_{23} - x_{11})^2 + (r_{sa}\sin\theta - z_{23} + z_{11})^2}}\right) + \tan^{-1}\left(\frac{r_{sa}\sin\theta - z_{23} + z_{11}}{r_{sa}\cos\theta + x_{23} - x_{11}}\right) \tag{35.29}$$

$$\eta_l = \sin^{-1}\left(\frac{-r_g + r_w}{\sqrt{(r_{sa}\cos\theta + x_{23} - x_{11})^2 + (r_{sa}\sin\theta - z_{23} + z_{11})^2}}\right) + \tan^{-1}\left(\frac{r_{sa}\sin\theta - z_{23} + z_{11}}{r_{sa}\cos\theta + x_{23} - x_{11}}\right) \tag{35.30}$$

expressing the angles η_u and η_l as functions of the swing-arm angle to the horizontal, θ, obtained by summing the nominal value with the displacement from nominal. For any swing-arm position, the coordinates of the points p_{gu}, p_{wu}, p_{gl}, and p_{wl} can therefore be specified.

The upper chain run is in tension if $(-r_g\varphi_g + r_w\varphi_w + \delta l_u)$ is positive, the first two terms accounting for the spinning of the sprockets and the third for the increase in separation of p_{gu} and p_{wu} due to suspension motion. Similarly, the lower run is in tension if $(r_g\varphi_g - r_w\varphi_w + \delta l_l - s_l)$ is positive, in which s_l represents the chain slack. If the upper chain run is in tension, it applies a force between the end points of the form $(-kx - cv)$ where k is the spring rate, c the damping coefficient, x is the extension, and v the extension speed. In such a case, the lower run force is zero and conversely.

For nonlinear simulation purposes, AutoSim's "ifthen" function is used to deal with the logic of determining whether the upper or lower chain run is in tension and therefore of deciding which forces to apply. For the corresponding linearization, the upper chain run only is used and the chain is considered to operate in tension or compression. It is implied that the chain slack is omitted from the linear model.

35.4.10 Speed and Steer Controllers

To maintain a desired forward speed profile, driving torque is applied to the gearbox sprocket and reacted on the main frame. The driving torque is not allowed to be negative. If braking is needed to maintain the desired speed, the front brake is used. The torque is produced by a proportional/integral control on the speed error with constant gains. The two gains required are noncritical. Effective values can easily be found by a few trials. If the integral gain is too large, oscillatory instability results. The target speed is provided as data in a table function, with time as the independent variable.

A steering feedback controller is also necessary to stabilize the machine in maneuvers in which it is not self-stable. In particular, stabilizing control allows the solution of the steady-turning equilibrium state problem by simply running a simulation to steady state. The controller devised is a proportional/integral/derivative (PID) feedback of motorcycle lean angle error to steering torque, with the lean angle target being set by an initial value and a constant rate of change. The target lean

angle must therefore be a ramp function of time. This would be easy to alter if it were considered restrictive. In fact, the PID-stabilizing controller can be used as part of a path-tracking control based on two-track vehicle practice [49]. A different approach to tracking control is considered in Section 35.7.

The steering control gains need to be speed adaptive, and they need choosing with considerable care to achieve effective stabilization. Especially difficult are cases involving very low or very high speed and high lean angles. Each of the three PID gains is linearly related to speed, as indicated by the relations

$$G_p = spg_0 + spg_1.u \tag{35.31}$$

$$G_i = sig_0 + sig_1.u \tag{35.32}$$

$$G_d = sdg_0 + sdg_1.u \tag{35.33}$$

corresponding to the control law

$$\tau = G_p(\phi - \phi_{ref}) + G_i \int_0^t (\phi - \phi_{ref})\, dt + G_d \dot{\phi} \tag{35.34}$$

where

u is the forward speed
τ is the steering control torque
ϕ is the lean angle
ϕ_{ref} is the target lean angle

Since the gains are critical only at very low and very high speeds, the constants in (3.21) through (3.23) can be chosen based on the extreme-speed behavior. Root locus analysis can conveniently be used to find values yielding good stability. Resulting gains as functions of speed for our basic motorcycle and rider, corresponding to coefficients $spg_0 = -24$, $spg_1 = -2.231$, $sig_0 = -163$, $sig_1 = 0.6154$, $sdg_0 = -48.5$, $sdg_1 = 0.6923$, are shown in Figure 35.11. These gains are tuned for 5 and 70 m/s in straight running.

35.4.11 Inputs

In addition to the feedback controls described earlier, braking torques between rear wheel and swing arm and between front wheel and lower forks can be applied in the model according to a time or velocity schedule. The brakes have been used in studies of deceleration dynamics [36]. Also, a

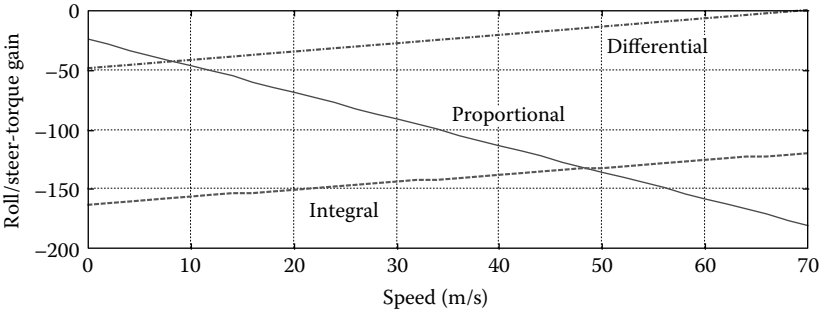


FIGURE 35.11 PID steering controller gains as functions of speed for standard motorcycle and rider.

simple longitudinal road profile is included, such that a length of road can be given a sinusoidal profile with a ramp function wavelength. The tire traverses the profiled road according to the point contact model described in Section 35.4.5, with the rear wheel seeing a version of the profile that is delayed relative to the front by the wheelbase travel time. It is implicit in this description that the road profiling is of wavelength substantially longer than the tire-to-ground contact. For short-wavelength irregularities, a more complex strategy would be needed [33].

As described in Section 35.4.10, the nonlinear simulation model is normally set up to find trim states by automatically tracking a demand for constant lean angle and slowly varying speed or for constant speed and slowly varying lean angle. The linear model, from substantially the same AutoSim source program, does not normally include the speed or steering controllers. To allow the computation of responses of the uncontrolled system to standard inputs in the linear world, those inputs to the linear model have to be defined.

Inputs included comprise driving torque to the gearbox sprocket for longitudinal control, steering torque between the steering frame and the rider upper body and rider upper body lean torque reacting on the rider's lower body for lateral control, and road profiling as a disturbance input [50].

35.4.12 Parameter Values

The full set of parameter values for the kind of model described is extensive and much work is involved in finding them. Most of the values belonging to the GSX-R1000 are listed in [47], with the remainder in [48], and they are not reproduced here. Figure 35.12 is a diagrammatic but geometrically accurate picture of the motorcycle. The figure shows eight masses each with diameter in proportion to mass represented, sprockets for the gearbox and rear wheel, and suspension details.

35.4.13 Equilibrium State Checking and Power Balancing

With the stabilizing controllers installed, the motorcycle can be run to equilibrium at any feasible speed and lean angle. To describe such an equilibrium state, force and moment balance equations can be set up, as was done in [47,50]. For health checking purposes, it is necessary that any equilibrium state yielded by the simulation model satisfies the force and moment balance conditions. As described in [47], the checking process also includes a power balance, whereby the engine power is

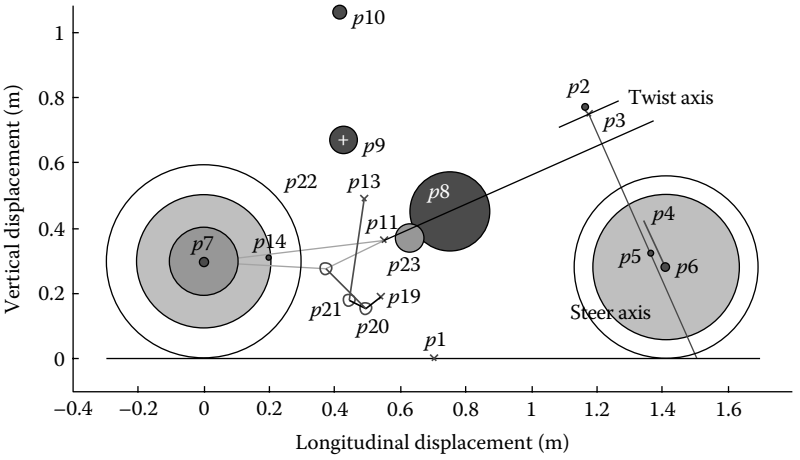


FIGURE 35.12 Pictorial representation of Suzuki GSX-R1000 machine in nominal state. Masses are shown in proportion to their magnitudes. The monoshock rear suspension and telescopic front forks are shown. The chain drive sprockets are also included at $p23$ and $p7$. $p1$ is at mid-wheelbase and is the aerodynamic force reference point.

shown to account precisely for the aerodynamic and tire losses. In steady turning, the force balance check is to ensure that the sum of the external forces is equal to the sum of the inertial and gravitational forces. The force error calculated is

$$\underline{F}_{error} = \sum_i \underline{F}_i + \sum_j m_j(\underline{g} - \underline{\omega}_j \times \underline{v}_j) \tag{35.35}$$

the first sum containing all the external forces, while the second deals with gravitational and centripetal effects. The external forces include (i) the aerodynamic lift and drag forces, (ii) the front and rear wheel normal loads, (iii) the tire longitudinal forces, and (iv) the tire side forces, including the driving force at the rear tire sufficient to maintain the steady speed. In the second term, m_j represents the mass of the jth body, $\underline{v}_j$ is the velocity of the body's mass center, $\underline{\omega}_j$ is the body's angular velocity vector, and g is the gravitational acceleration vector.

In much the same way, the following moment error should be zero:

$$\underline{M}_{error} = \sum_i \underline{l}_i \times \underline{F}_i + \sum_j \{\underline{l}_j \times m_j(\underline{g} - \underline{\omega}_j \times \underline{v}_j) - \underline{\omega}_j \times \underline{H}_j\} + \sum_k \underline{M}_k \tag{35.36}$$

where

l_i and l_j are moment arm vectors referred to the rear wheel contact point

H_j is the moment of momentum of body j about its mass center

The first sum accounts for the moments generated by the external forces listed earlier, while the second contains a part treating gravitational moments and moments of inertial forces on the body mass centers and a part accounting for the rate of change of moment of momentum of each body, with respect to its mass center. The third summation deals with aerodynamic pitching and tire aligning moments.

Each of the terms $\underline{\omega} \times \underline{H}$ is calculated as $\underline{\omega} \times \left(H_x \underline{i} + H_y \underline{j} + H_z \underline{k}\right)$, with $\underline{H}$ having components H_x, H_y, and H_z in directions denoted by the unit vectors $\underline{i}$, $\underline{j}$, and $\underline{k}$, which must be chosen so that the moment of momentum components is invariant, when the motorcycle is in a steady turn. For all the non-spinning bodies, the body reference axes satisfy this requirement. For the wheels, the parent body's reference system needs to be used and a moment of momentum term for the spin added on. For the most general case applicable here, in which I_{xy} and I_{yz} are zero but I_{xz} is nonzero [51], noting a change of sign of products of inertia is compared with the reference because [51] and AutoSim use opposite conventions:

$$H_x = I_{xx}\omega_x + I_{xz}\omega_z \tag{35.37}$$

$$H_y = I_{yy}\omega_y \tag{35.38}$$

$$H_z = I_{xz}\omega_x + I_{zz}\omega_z \tag{35.39}$$

in which ω_x, ω_y, and ω_z are the components of $\underline{\omega}$ in the $\underline{i}$, $\underline{j}$, and $\underline{k}$ directions.

Thus, for the main body, this second component of rate of change of moment of momentum, that about the mass center, is of the form

$$\underline{\omega}_{main} \times \{(I_{mainx}\omega_x + I_{mainxz}\omega_z).\underline{i}_{main} + I_{mainy}\omega_y.\underline{j}_{main} + (I_{mainxz}\omega_x + I_{mainz}\omega_z).\underline{k}_{main}\} \tag{35.40}$$

Here, $\underline{i}_{main}$, $\underline{j}_{main}$, and $\underline{k}_{main}$ denote the unit vectors $\underline{i}$, $\underline{j}$, and $\underline{k}$ for the main body. To deal with the rear wheel, its diametral inertia is added to the corresponding terms belonging to the swing

arm, its parent body, as if it were part of the swing arm. The spin is accounted for by a term $\underline{\omega}_{swingarm} \times I_{rwy}\omega_y.[rwy]$ and similarly for the front wheel, for which the lower fork body is the parent.

The power error is given by

$$P_{error} = \tau.\omega_{spin} + \sum_i \underline{F}_i.\underline{v}_i + \sum_k \underline{M}_k.\underline{\omega}_k \tag{35.41}$$

where

τ is the gearbox sprocket driving torque

ω_{spin} is its spin velocity relative to the main frame

$\underline{v}_i$ is the velocity of the point of application of force, $\underline{F}_i$

$\underline{\omega}_k$ is the absolute angular velocity of the body to which moment, $\underline{M}_k$, is applied

Describing the power associated with tire forces requires care also. The velocity involved is that of the tire tread base material [50,51], already calculated in connection with finding the slip ratio and the slip angle.

These checks on any equilibrium state are substantially independent of the full equations of motion on which the simulation model depends, and it is reassuring that they are always satisfied to a high level of accuracy.

35.5 MOTORCYCLE BEHAVIOR

35.5.1 Equilibrium "Trim" States

If the allowed motions include only small perturbations from straight running at constant speed, longitudinal and vertical motions are decoupled from the lateral motions by virtue of the motorcycle symmetry, and much understanding of the steering behavior can be obtained at this level, the lateral and longitudinal/vertical problems being considered separately.

When the motorcycle is run from low speed to high speed at very low acceleration levels, it passes through a sequence of quasi-equilibrium states. As the speed builds, the drive torque has to increase as the aerodynamic drag, lift, and pitching moment increase according to the square of the speed. The machine attitude adjusts correspondingly, as can be seen in Figures 35.13 and 35.14. Note especially the suspension system deflections increasing systematically with speed, contrasting with the assumptions of [52].

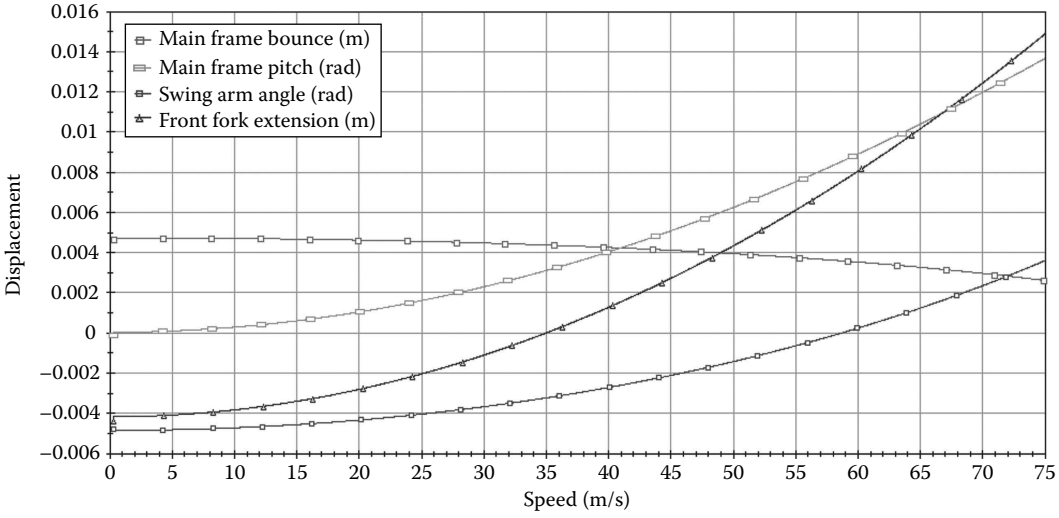


FIGURE 35.13 Straight-running attitude results showing how aerodynamic forces grow with speed and influence the posture of the motorcycle. Positive directions are downward for bounce, nose up for pitch, and rebound for the swing arm and the front forks.

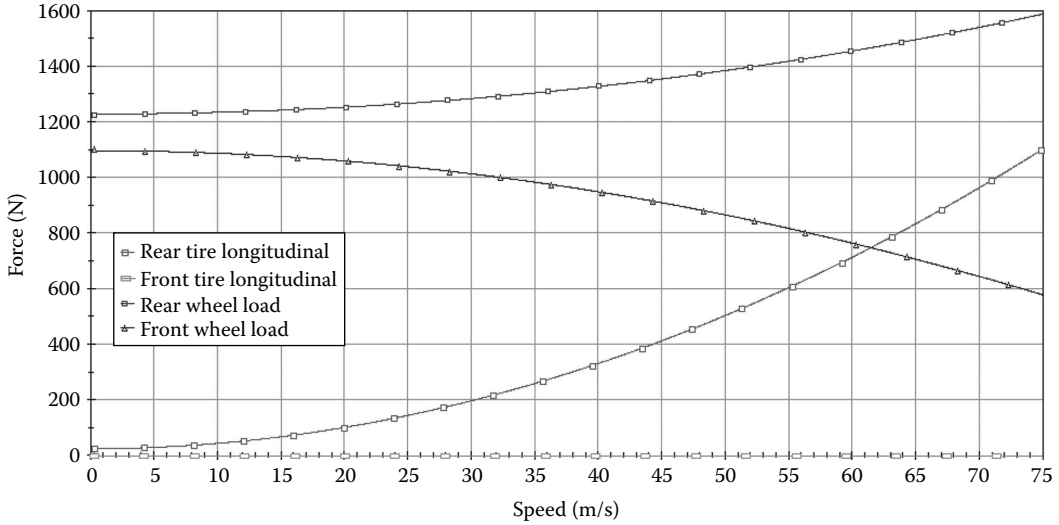


FIGURE 35.14 Straight-running attitude results showing how tire forces change with speed. Driving force is positive. In this example, the driving force at the rear tire is about 70% of the load on the tire at 75 m/s.

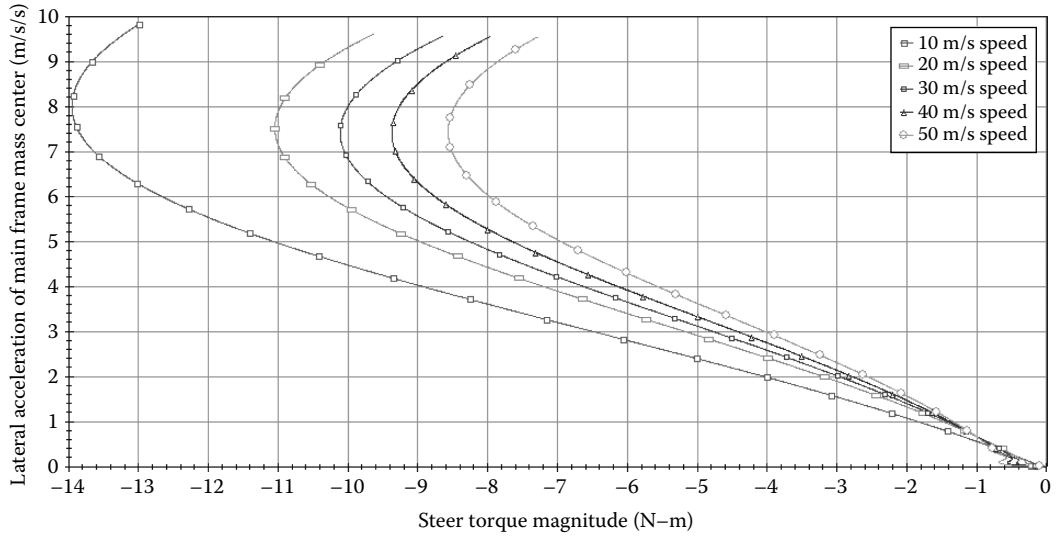


FIGURE 35.15 Steady-state lateral acceleration responses to steering torque input at a series of fixed speeds with slowly increasing lean angle. Infinite slope of a curve, occurring between 7.3 and 8 m/s/s lateral acceleration in these cases, indicates a possible point of transition from static stability to static instability.

In cornering, the speed is conveniently fixed, and the lean angle is ramped up to somewhere near the limit of the available performance space. In these cases, handling diagrams showing lateral acceleration of the main frame mass center against either steer torque (free control) or steer angle (fixed control) can be plotted. These diagrams need to be interpreted strictly in accordance with the theoretical background exposed in Chapter 2. Examples for speeds of 10, 20, 30, 40, and 50 m/s and lean angles up to 0.9 rad are shown in Figures 35.15 and 35.16. If the steer torque is regarded as the input, the steer angle is itself a response, but when the fixed-control stability properties are suitable, the rider does have the option of using steer angle as the control input, utilizing his or her own internal control capabilities to produce the required steering displacements.

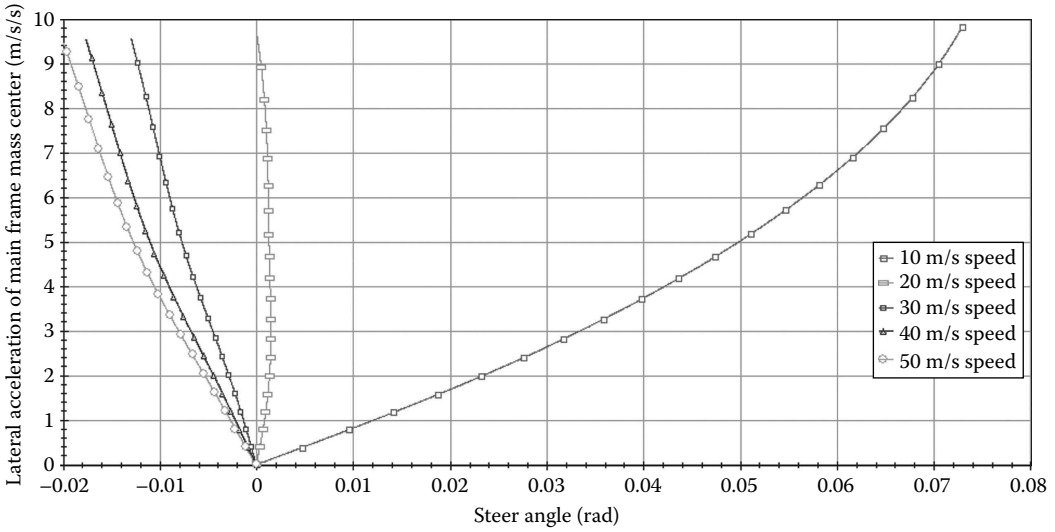


FIGURE 35.16 Steady-state lateral acceleration responses to steering angle input at a series of fixed speeds with slowly increasing lean angle.

The full spectrum of results for each of these cornering cases is complex and all the details are not shown. However, they are available to the linearization process through result files that can be read into MATLAB and utilized to evaluate the matrix elements of the linear model. The passive rider lower and upper bodies displace roughly in proportion to the main frame roll angle. The lateral displacement of the lower body is always less than 3 mm, while the roll displacement of the upper body is approximately 10% of the main roll angle. The wheel loads remain much the same throughout, but the suspensions compress, as the tire side forces grow. The front tire drag increases to about 100 N as the lateral acceleration builds toward 10 m/s/s.

35.5.2 Stability for Small Perturbations from Straight-Running Trim States

Suppose we use our linear model in straight running to calculate first the fixed-control and secondly the free-control stability properties of the motorcycle/rider system. Each process (in MATLAB) involves reading the equilibrium states for the running condition of interest, evaluating the elements of the A matrix (Equation 35.3), finding the eigenvalues and plotting those that lie within an interesting range, and moving to the next running condition to repeat the computations until the so-called root locus [35] is complete. Results can be plotted either as real parts of roots against speed or as imaginary parts against real parts with speed as the varying parameter.

Results are shown in Figures 35.17 and 35.18. In fixed control (steering system locked in the straight-ahead position), the machine system is unstable in roll at low and medium speeds, with a positive real eigenvalue of about 3.75 at the lowest speeds that slowly decreases as the speed increases. The motion belonging to this mode is akin to the divergence from the upright configuration of an inverted pendulum. The mode stabilizes at about 23 m/s. Then, surprisingly, the fixed-control motorcycle and rider are stable until an oscillatory mode with frequency approaching 3 Hz becomes unstable at about 65 m/s. Rider stabilization of the low-speed motion by steering position adjustments should not be impossible but would not be easy. The time constant of the divergent rolling motion is that of a uniform rod of approximate length 1.2 m balancing on its pointed end. However, at speeds between 23 and 65 m/s, this fixed-control motorcycle is self-stabilizing, and it appears to be entirely possible for a rider to use primarily steer angle control in this speed range. The autostabilization must arise primarily from the development of lateral forces by the tires when they run at a camber angle. As the motorcycle leans, the wheel camber and the forces generated apply a rolling moment to the system to counter the leaning.

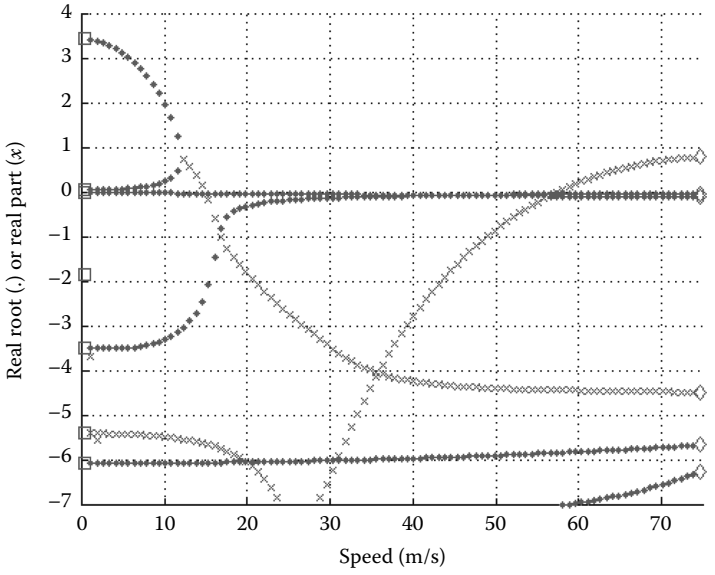


FIGURE 35.17 Eigenvalue real parts as functions of speed in the range 0.5–75 m/s for straight-running standard motorcycle in fixed control. The standard model is converted effectively to fixed control by multiplying the standard steering damper coefficient by 100. Remarkably, the fixed-control motorcycle self-stabilizes in straight running between about 14 and 57 m/s.

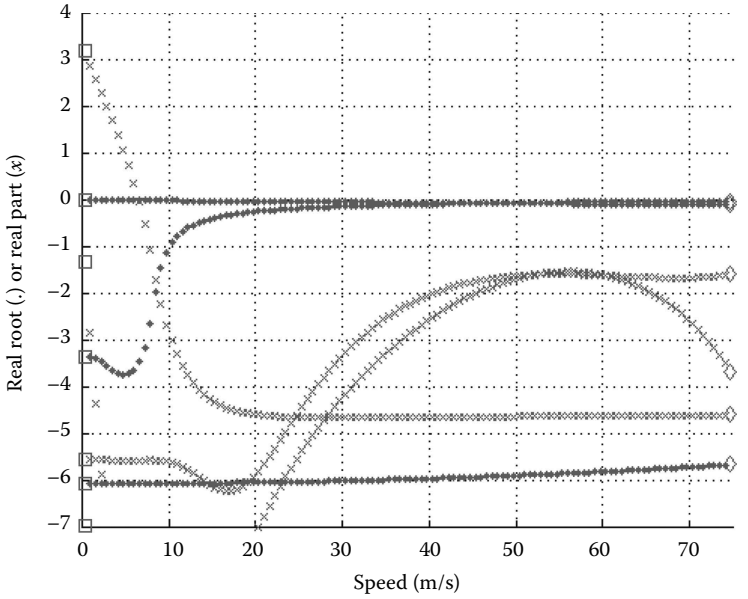


FIGURE 35.18 Eigenvalue real parts as functions of speed in the range 0.5–75 m/s for straight-running standard motorcycle in free control. The free-control motorcycle self-stabilizes in straight running at about 6 m/s.

The locus is shown again in the alternative form in Figure 35.19, along with the results of halving and doubling the frame torsional stiffness. In free control, at very low speeds, there are two positive real eigenvalues. One of these can be associated loosely with the whole machine rolling over or capsizing, as with fixed control, while the other is related to the steering system rotating from the straight-ahead nominal position to full lock under the influence of gravity. With a very small increase in speed, these two real roots become a complex conjugate pair describing an oscillatory

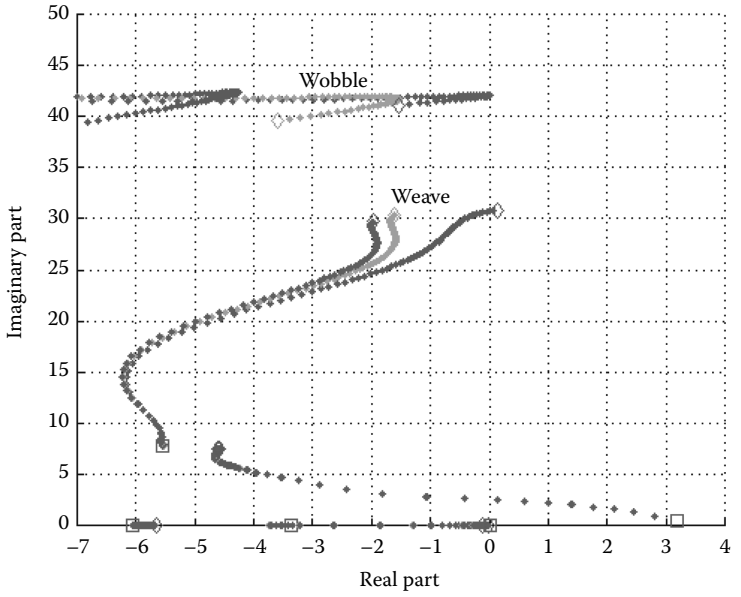


FIGURE 35.19 Root loci for the speed range 0.4–75 m/s for the standard motorcycle with frame torsional stiffness variations. The red curves show the results of halving the stiffness, while the magenta curves show the consequences of doubling the stiffness. The low-speed end of each locus is indicated by a square plot symbol, while the high-speed end is shown by a diamond.

motion of very low frequency, becoming stable at about 6 m/s (see Figure 35.18). This autostabilization of the system at relatively low speed makes the free-control properties of the motorcycle much more attractive to the rider than the fixed-control ones. It can be safely concluded that the rider's main control actuation at low speeds is by steering torque and it is very important that the steering system is free to turn at these low speeds, where the fixed-control behavior is particularly poor.

It has been suggested that the wobble mode can be viewed as involving oscillations of an isolated castor [29,53], the steering axis being fixed and the rear frame properties having no influence on the stability. The idea has been tested in practice by the Pirelli Company, who applied their "pendulum" method [54] to motorcycle front fork assemblies, finding a medium-speed wobble instability to occur regularly. The idea can be tested theoretically in the following way. The stability of the front steering frame on its own can be found by adding stiff lateral restraining springs and dampers to the point $p2$ in Figure 35.12 and to another point on the steering axis near the ground, sufficient to prevent the steer axis from moving laterally. Root loci for the restrained system are then found in the conventional way, with attention only to the wobble-mode locus (Figure 35.20). The medium-speed wobble instability is clear, showing that the location of the steering axis by the rear frame is important to the stability of the front frame and confirming that the wobble mode cannot properly be considered in isolation from the remainder of the system.

When the motorcycle frame is relatively flexible in torsion, as has commonly been the case historically, the wobble mode is minimally damped at medium speed (Figure 35.19), as with the steering axis fixed. The characteristic behavior, in a problematic case, is for the steering assembly to go into sustained low-amplitude oscillations at or near a particular speed, as long as the rider does not impede the motions of the handlebars. The extent to which the rider might control a wobble oscillation by deliberate intervention is considered in [25]. Depending on the frame compliance, the worst speed in terms of the modal damping may be that for which the front wheel rotation frequency is equal to the wobble frequency, at around 6.6 Hz, which is better avoided by design. In such a case, any imperfections in the front wheel or tire (out of round, unbalance, etc.) will contribute forcing, and the resonant response will tend to be excessive. The possible coincidence of forcing and natural frequencies is illustrated in Figure 35.21, computed for the standard machine.

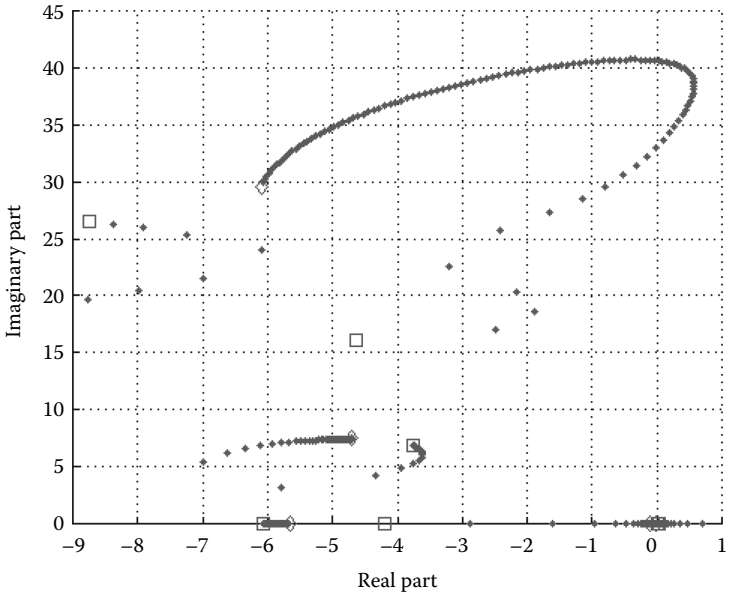


FIGURE 35.20 Root locus for standard machine in straight running with lateral stiffness and damping restraints added to the steer axis, to prevent its lateral motion. As has been shown previously [55], the wobble mode is unstable at moderate speeds, if the steer axis has no lateral mobility.

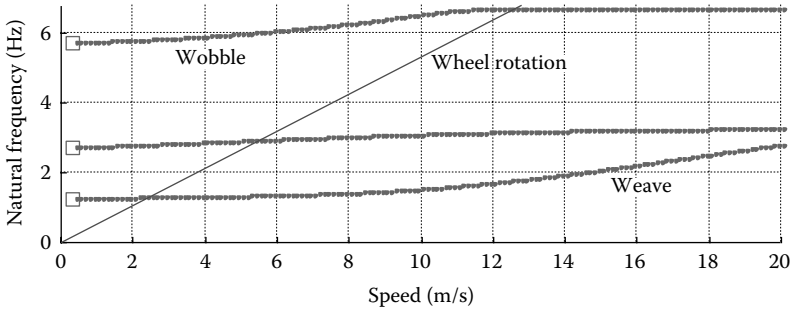


FIGURE 35.21 Typical wobble- and weave-mode natural frequencies and front wheel rotation frequency, as functions of speed, illustrating the possibility of resonant forcing of the wobble mode at moderate speed, in the case of a compliant-framed machine.

When the frame is stiff in torsion, the wobble mode is least damped at higher speeds (Figure 35.19), and it can be expected that wobbles will occur at higher speeds also. A steering damper is effective in stabilizing the wobble oscillations, but it carries a cost in terms of destabilizing the weave (Figure 35.22) and increasing the muscular effort required from the rider for steering control. There is evidence that riders have little capacity to damp out a high-speed wobble, once it becomes established to the level of a few degrees of handlebar amplitude [25]. It is dangerous due to this feature. Recent work has pointed to the possibility of a passive mechanical network being an advantageous and practical alternative to the steering damper, providing the beneficial influence of the damper on the wobble mode but also improving the weave and not implying heavy steering [56,57] (see Section 35.5.9). In the kind of network proposed, the transfer properties of the network can be shaped with frequency, such that the desirable influence of a steering damper on the wobble mode can be retained, while the deterioration of the weave mode can be avoided.

The importance of including rider compliances, tire sideslipping, and tire relaxation lags in the modeling is illustrated in Figures 35.23 through 35.25. In the first of these cases, the standard rider stiffnesses

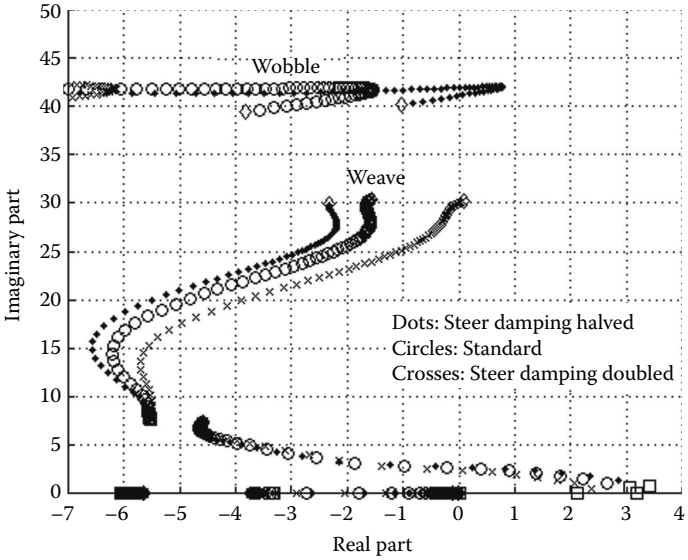


FIGURE 35.22 Root loci for standard motorcycle with steering damper coefficient halved (dots), normal (circles) and doubled (crosses). The steering damper is effective in stabilizing the wobble mode but de-stabilizes the weave mode. The low-speed end of each locus is indicated by a square plot symbol, while the high-speed end is shown by a diamond.

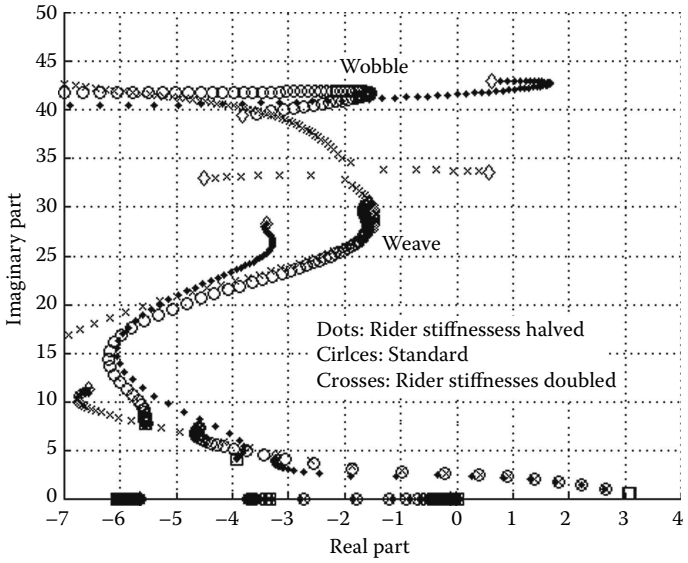


FIGURE 35.23 Root loci with rider stiffness and damping parameters varied. Both stiffnesses are halved (dots) and doubled (crosses) with respect to the nominal settings (circles). The corresponding damping coefficients are varied by a factor $\sqrt{2}$ each time. The low-speed end of each locus is indicated by a square plot symbol, while the high-speed end is shown by a diamond.

are halved and doubled, while the nominal settings, showing the same results as in Figure 35.19, are included for reference purposes. The damping coefficients corresponding to the stiffnesses altered are varied according to the square root of the stiffness adjustment factor. In the second case, both front and rear tire cornering stiffnesses are halved and doubled in the same way, while in the third case, both tire relaxation lengths are halved and doubled similarly. The stability properties of the motorcycle are

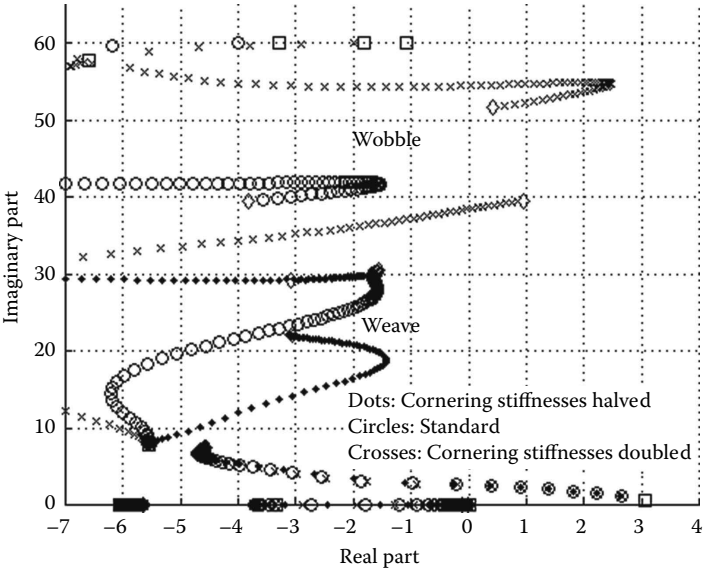


FIGURE 35.24 Root loci with tire cornering stiffnesses halved (dots), normal (circles), and doubled (crosses). The low-speed end of each locus is indicated by a square plot symbol, while the high-speed end is shown by a diamond.

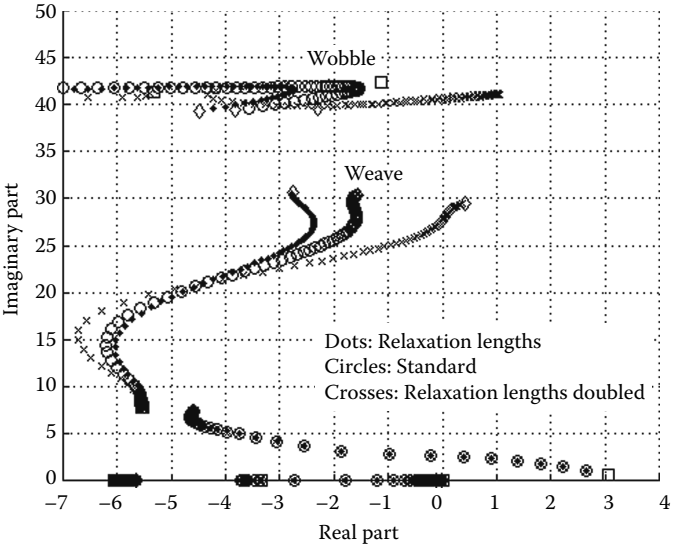


FIGURE 35.25 Root loci with tire relaxation lengths halved (dots), normal (circles), and doubled (crosses). The low-speed end of each locus is indicated by a square plot symbol, while the high-speed end is shown by a diamond.

changed grossly by such large variations in parameter values, and it is clear that the features scrutinized, namely, frame torsion and rider compliances, tire sideslipping and tire relaxation lags, must be included in models of contemporary machines to make good predictions of behavior over the full speed range.

35.5.3 Normal Modes for Straight-Running Trim States

Mode shapes are given by appropriate eigenvector columns. Examples for the wobble mode are shown in Figure 35.26 and for the weave mode in Figure 35.27. In these diagrams, "lat" means the lateral velocity of the main frame reference point, nominally coinciding with the rear tire/ground contact point; "yaw"

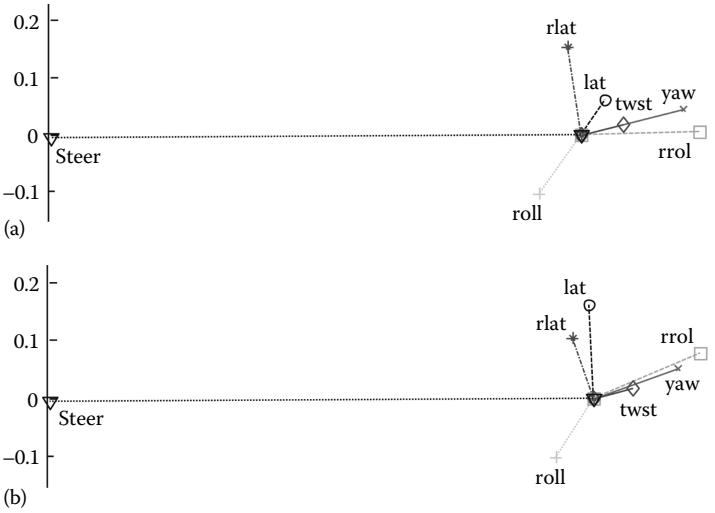


FIGURE 35.26 Wobble-mode eigenvectors for 20 (a) and 50 m/s (b). Translational velocities in m/s and angular velocities in rad/s. Corresponding eigenvalues are –7.142 ± 41.743i and –1.686 ± 41.880i.

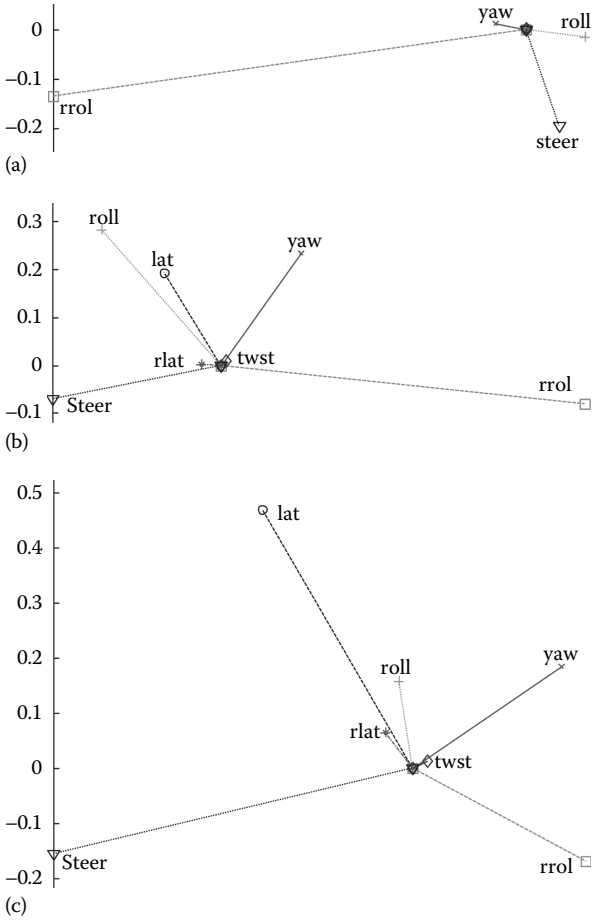


FIGURE 35.27 Weave-mode shapes for 5 (a), 20 (b), and 50 m/s (c). Translational velocities in m/s and angular velocities in rad/s. Corresponding eigenvalues are –5.575 ± 8.18i, –5.866 ± 17.35i, and –1.622 ± 27.35i.

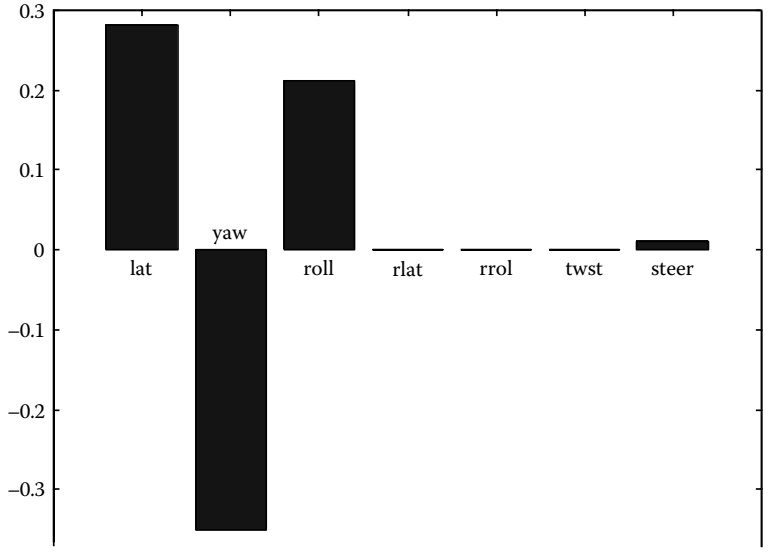


FIGURE 35.28 Capsize mode participant motions at 20 m/s. Translational speeds in m/s and rotational speeds in rad/s.

and "roll" are for the main frame; "rlat" means the lateral velocity of the rider lower body relative to the main frame; "rrol" means the rider upper body relative roll velocity; "twst" is the frame twisting velocity; and "steer" is the steer velocity relative to the main frame. The wobble mode mainly involves contrarotation of the steering system and the main frame, with significant participation from roll, rider motions, and frame twisting. The shape does not change much over the speed range. The weave mode is more variable with speed and is harder to characterize. There is always substantial rider participation, especially from rolling of the upper body. This rider participation is, of course, a consequence of the adoption of the rider model from [13], derived from excitation of a number of riders in pure roll. To the extent that the modal motions are different from this, the direct acceptance of the results is of questionable validity. The best way to represent a generic rider remains an open issue. Animation of the modes is a very useful aid to obtaining some appreciation of them (http://www.ic.ac.uk/controlandpower/motorcycles).

The capsize mode at 20 m/s, having real eigenvector elements, is depicted in Figure 35.28. The motion involves translating to the right (say), rolling to the right, and steering to the right, while the motorcycle yaws to the left. Rider and frame twisting contributions are negligible.

35.5.4 Frequency Responses in the Neighborhood of Straight Running

From the small-perturbation linear model, it is straightforward to calculate frequency responses to defined excitation, but it is not easy to interpret the results of such calculations. It was demonstrated in [3] that motorcycles are not as good as cars in responding to a sudden need to maneuver and that the skill of the rider is a substantial issue in avoiding obstacles (see [30] also). It is clear that maneuverability has some importance but it is not so clear how to define it objectively. One aspect of the problem that can be enlightened by calculation is the relative responsiveness of the motorcycle to steer torque and to rider lean torque forcing, an issue studied experimentally in [8]. Let us choose to compute yaw rate responses to steer torque or rider lean torque input. Since motorcycles only rarely operate with significant sideslipping, the yaw rate response is considered a reliable indicator of the lateral motions and the path followed.

Yaw rate responses to steer torque input for the standard motorcycle at speeds of 10, 30, 50, and 70 m/s are shown in Figure 35.29, and corresponding responses to rider upper body torque input are shown in Figure 35.30. The responses at the low frequencies of importance for rider control reduce as the frequency rises and are much reduced for higher speeds. Weave and wobble resonances

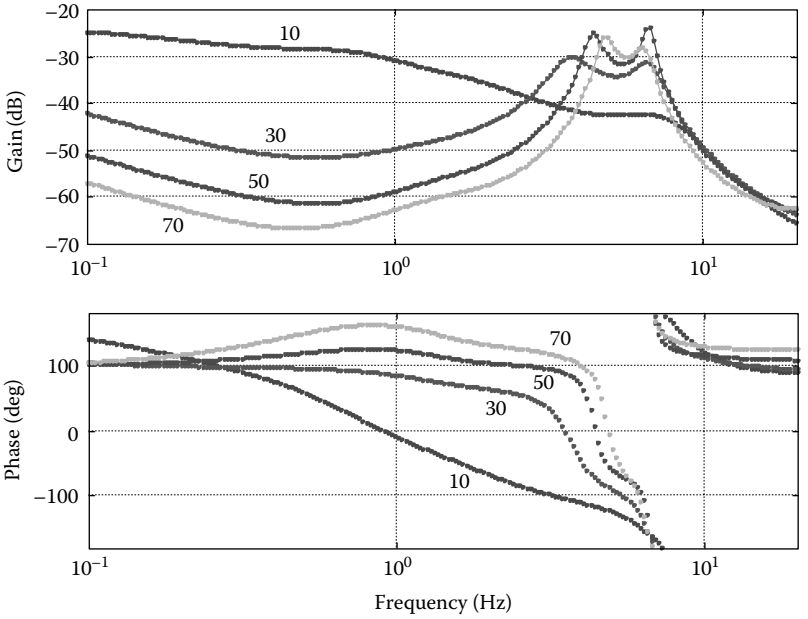


FIGURE 35.29 Yaw rate frequency responses to steer torque input for standard machine in straight running re 1 (rad/s)/(N m) for speeds 10, 30, 50, and 70 m/s.

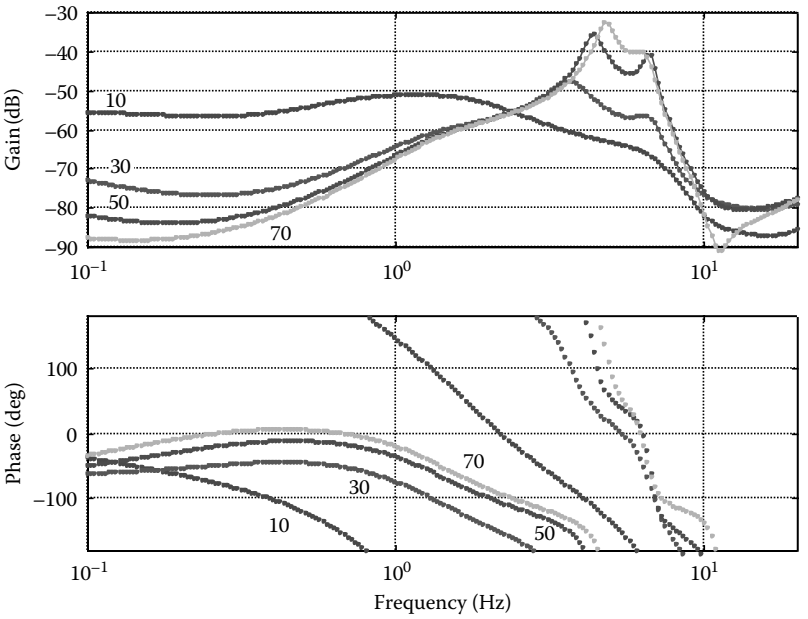


FIGURE 35.30 Yaw rate frequency responses to upper body lean torque input for standard machine in straight running re 1 (rad/s)/(N m) for speeds 10, 30, 50, and 70 m/s.

show clearly for the higher speeds. Responses to rider body torque are much less than those to steer torque, confirming the idea that steer torque control is completely dominant over body torque control. If the main frame roll inertia or the main frame yaw inertia is doubled, and the results of Figure 35.29 are regenerated, only very small changes occur. These results are not shown, since they are so similar to those of Figure 35.29. Responses to wheel rotation forcing, with various types of wheel or tire imperfection, are shown in [58], which results are not repeated here.

35.5.5 Stability in Cornering

When motorcycle motions involve large excursions from the nominal laterally symmetric state, out-of-plane movements, lateral, yaw, roll, rider lean, frame twist, and steer, become interactive at first-order level with in-plane movements, bounce, pitch, and wheel hop. Trim states from the simulation runs used to yield Figures 35.15 and 35.16 are chosen now to illustrate the stability properties in cornering. The simulations involve fixing the speed and slowly ramping up the roll angle to near limit states, so the varied parameter in the root loci is the motorcycle roll angle or, equivalently, the lateral acceleration of the main frame mass center.

With free steering, from Figure 35.15, we might anticipate a transition from statically stable to statically unstable at a lateral acceleration level of 7.3–8 m/s/s for any of the five speeds. Examples are given in Figure 35.31, for 10 m/s speed, and Figure 35.32, for 30 m/s speed, with eigenvalue real parts plotted as functions of lateral acceleration, and Figure 35.33, for 50 m/s speed, which is a root locus plot, with lateral acceleration the swept parameter. Figures 35.31 and 35.32 show the expected appearance of positive real eigenvalues at high lateral accelerations. In each case, the stability transition is represented by the real part of a complex conjugate pair of eigenvalues changing sign, but the imaginary part is less than 0.03. The appearance of the complex pair near to the stability boundary is thought to be due to numerical imprecision in finding the eigenvalues. Figure 35.33 shows only a small influence of cornering on weave and wobble roots at 50 m/s speed. Corresponding root loci for other speeds show the same feature but they are not included.

Fixed steering is represented approximately by multiplying the standard steering damper coefficient by 100. From Figure 35.17, we should expect static instability for 10 m/s forward speed and static stability for 40 m/s speed at small lean angles. Figure 35.34, for 10 m/s speed, demonstrates static instability through the lateral acceleration range, while Figure 35.35, for 40 m/s, shows no positive real roots until the lateral acceleration reaches about 7.6 m/s/s.

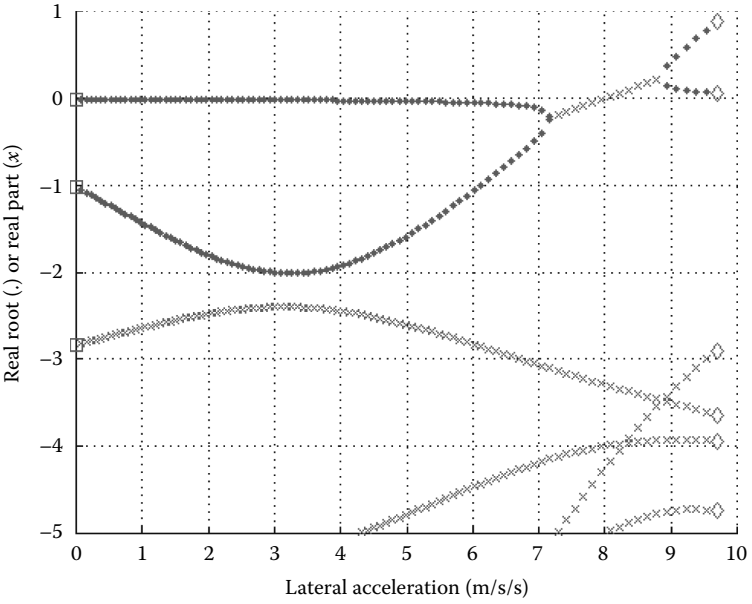


FIGURE 35.31 Real parts of characteristic roots for standard motorcycle in free control at a constant speed of 10 m/s through the full range of lateral accelerations. Dots show real roots, while crosses indicate real parts of complex conjugate roots. The crossing of the stability boundary occurs at 8 m/s/s lateral acceleration, as might be expected from Figure 35.15. The complex roots in this region are thought to be due to computational imprecision.

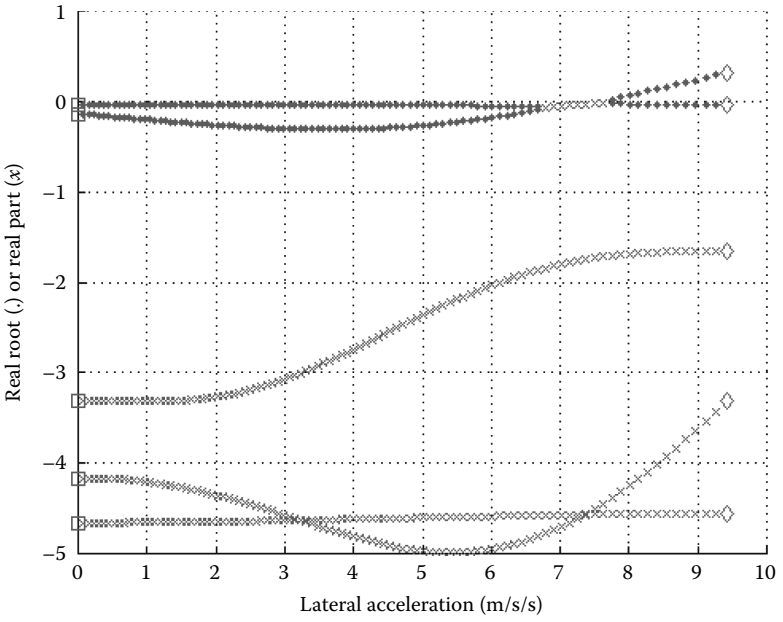


FIGURE 35.32 Real parts of characteristic roots for standard motorcycle in free control at a constant speed of 30 m/s through the full range of lateral accelerations. Dots show real roots, while crosses indicate real parts of complex conjugate roots. The stability boundary is crossed at 7.7 m/s/s lateral acceleration, again in line with Figure 35.15. Again, the complex roots near the boundary have very small imaginary parts.

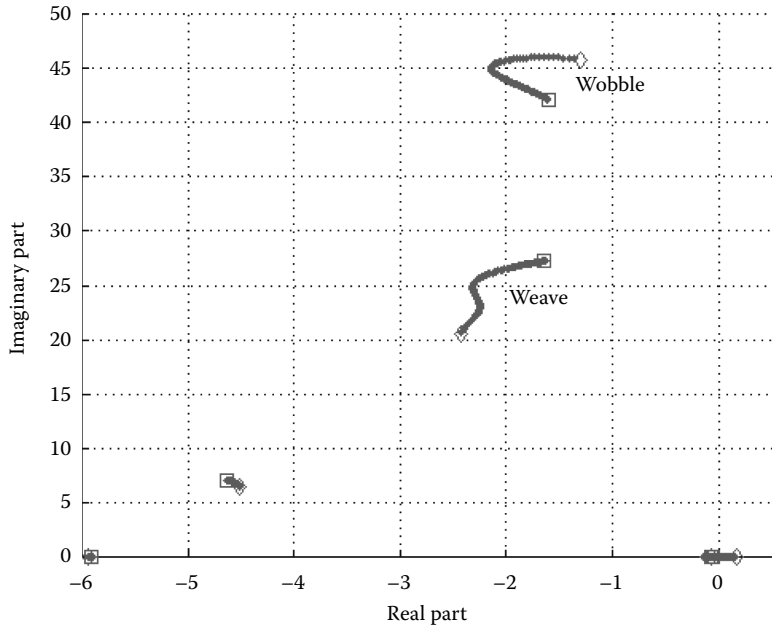


FIGURE 35.33 Root locus plot for a fixed speed of 50 m/s and lateral accelerations from 0 to 9.5 m/s/s. Squares are for zero lateral acceleration and diamonds for 9.5 m/s/s. The influence of cornering on the stability of the oscillatory modes is modest, the weave mode tending to improve with turning vigor.

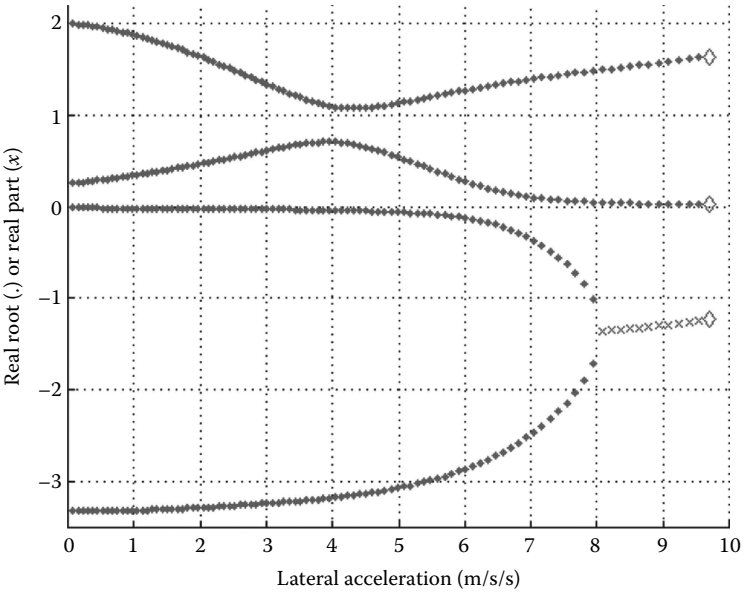


FIGURE 35.34 Real parts of characteristic roots for standard motorcycle in fixed control (steer damping*100) at a constant speed of 10 m/s through the full range of lateral accelerations. Dots show real roots, while crosses indicate real parts of complex conjugate roots. The system is statically unstable at this speed throughout the range of lateral accelerations.

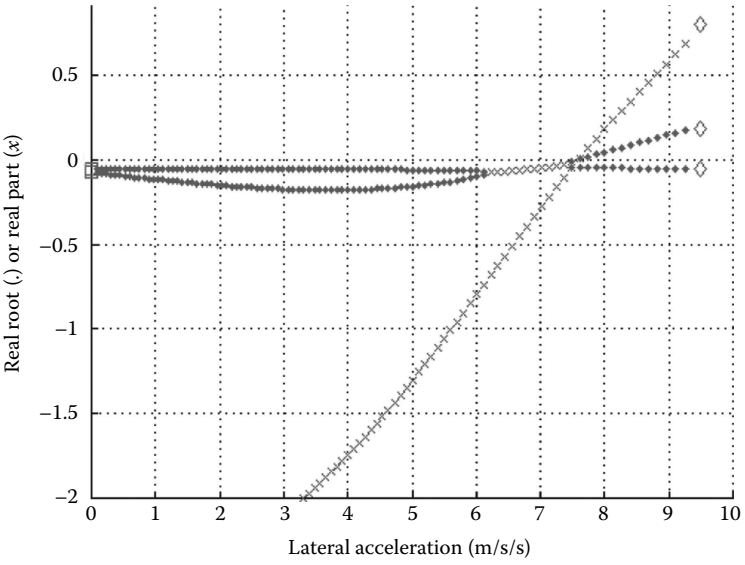


FIGURE 35.35 Real parts of characteristic roots for standard motorcycle in fixed control (steer damping *100) at a constant speed of 40 m/s through the full range of lateral accelerations. Dots show real roots, while crosses indicate real parts of complex conjugate roots. The motion becomes unstable at about 7.6 m/s/s lateral acceleration.

Root loci for the standard motorcycle for a constant roll angle of 20° and a speed range of 5–70 m/s are shown in Figure 35.36. The least stable speed for the weave mode is about 47 m/s and for the wobble mode is about 57 m/s. These running conditions will be examined from the point of view of road undulation forcing in Section 35.5.8.

The stability properties show that at low speeds, free steering of the motorcycle is essential to its controllability with reasonable effort. For higher speeds, fixed control is feasible, and it is likely that

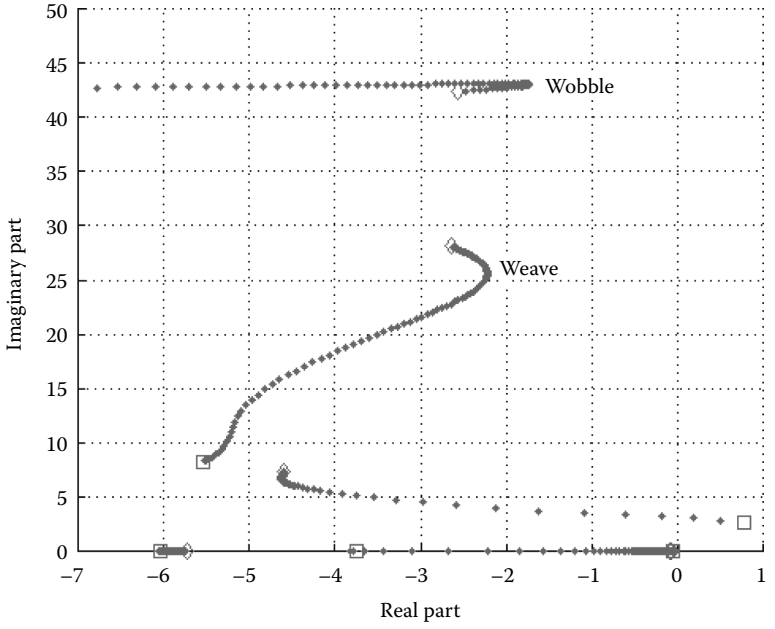


FIGURE 35.36 Standard machine root loci for speed 5–70 m/s and lean angle of 20°. Minimum damping of the weave mode occurs at about 47 m/s, while that for the wobble mode occurs around 57 m/s. Squares are for 5 m/s, while diamonds indicate 70 m/s speed.

real riders choose some convenient combination of free and fixed control, depending on machine and speed. At higher speeds, the fixed-control machine becomes oscillatory before the free-control machine does. At high speeds, the rider will again choose primarily free control. It is the stability properties in free control that matter most, since this is always the fallback option.

35.5.6 Cornering Modes

Due to the cross-coupling of in-plane and out-of-plane motions in cornering, the oscillatory modes have a more complex composition when the small perturbations occur about a turning trim state. The eigenvectors, showing the mode shapes, are just as easy to calculate as in straight running, but they are even more difficult to visualize. Again animation is potentially valuable (see (http://www.ic.ac.uk/controlandpower/motorcycles)). Cornering weave is illustrated in Figure 35.37 for 50 m/s speed and very high lateral acceleration, while cornering wobble is shown in Figure 35.38 for the same condition.

35.5.7 Road Undulation Forcing

Lightly damped oscillatory modes of motion have the potential to resonate in response to suitable regular forcing, as already mentioned in Section 35.5.2 in relation to wheel rotation forcing. Other possible sources of regular forcing are aerodynamic vortex shedding [22,59] and road irregularities [50]. It is believed that a whole class of motorcycle accidents is connected with road undulation forcing.

When a motorcycle is traveling substantially along a straight path, the coupling between road undulations and lateral responses is very weak and problems should not arise, but in cornering, the strength of coupling grows with the roll angle and sustained forcing along the curved path of the machine may lead to dangerously high response amplitudes. Resonant responses may be large for special combinations of machine speed, lean angle, and road wavelength. The speed and lean

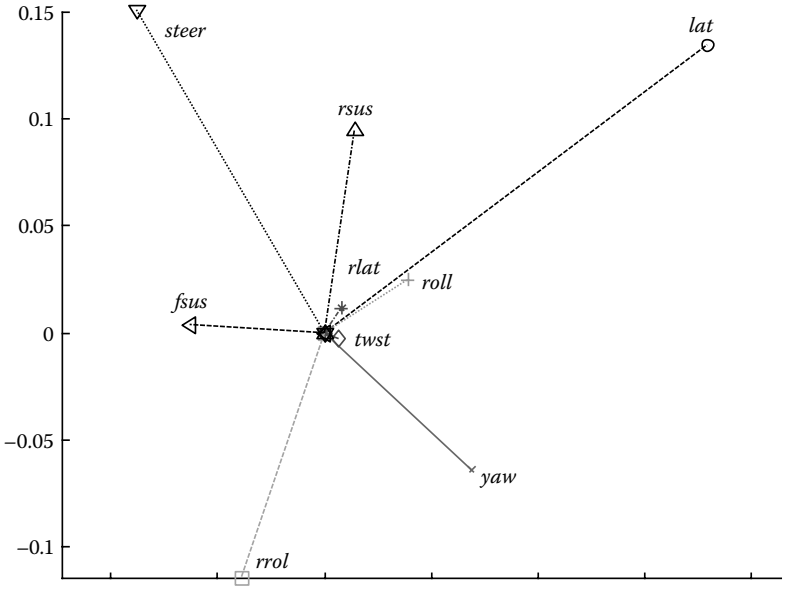


FIGURE 35.37 Weave-mode eigenvector for 50 m/s speed and 9.54 m/s/s lateral acceleration. The corresponding eigenvalue pair is –2.447 ± 20.39i. Translations are in m/s and rotations in rad/s. All the variables illustrated are velocities. "rsus" is the swing-arm angular velocity, while "fsus" is the front fork extension velocity. "rrol" and "rlat" are rider roll and lateral velocities.

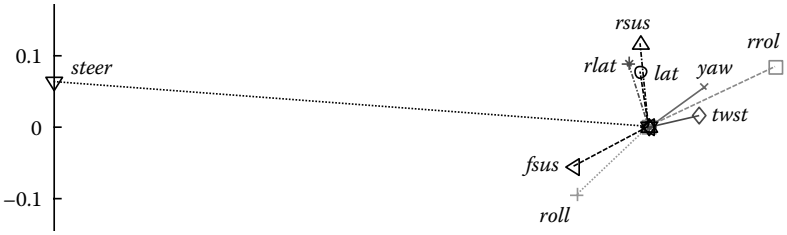


FIGURE 35.38 Wobble-mode eigenvector for 50 m/s speed and 9.54 m/s/s lateral acceleration. The corresponding eigenvalue pair is –1.278 ± 45.75i. Translations are in m/s and rotations in rad/s. All the variables illustrated are velocities. "rsus" is the swing-arm angular velocity, while "fsus" is the front fork extension velocity.

angle determine the natural frequencies, the modal damping factors, and the road forcing to vibration mode coupling strength. The wavelength governs the matching of forcing frequency to natural frequency. A detailed study of a touring machine from the early 1980s can be found in [50]. The connections with recorded accidents are also made there.

The calculation of responses to sustained road forcing was described earlier in Chapter 2 (see (2.10) in particular). In Figure 35.39, the results of such calculations for our standard machine at 20° roll angle and 57 m/s speed, where the wobble-mode damping is least, are shown. Similar results for 47 m/s, for which the weave mode is least damped, are shown in Figure 35.40. In each case, the two resonances are clear, and the 30 dB peak of the wobble mode at 57 m/s indicates that the road displacement amplitude necessary to cause a lock to lock steering response is only 16 mm. The road undulations do not need to be large to create a problem, as long as they are regular and are tuned to the machine properties at the relevant running conditions. In contrast to [50], forcing of the weave mode from the rear wheel is dominant with our standard machine, while the front wheel input is only a little stronger than that from the rear in respect of the wobble mode.

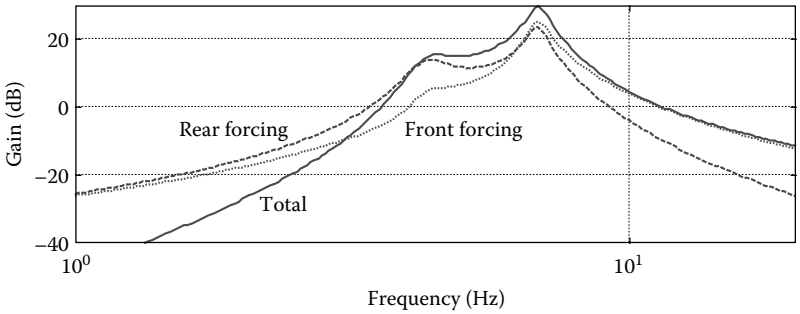


FIGURE 35.39 Steer angle frequency responses of standard bike at 57 m/s speed and 20° roll angle to road forcing in m (0 dB response represents 1 rad). In contrast to the case studied in [50], the rear wheel excitation is dominant in relation to the weave mode here.

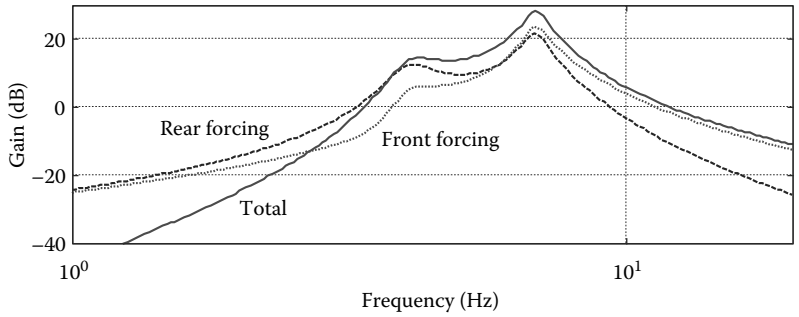


FIGURE 35.40 Steer angle frequency responses of standard bike at 47 m/s speed and 20° roll angle to road forcing in m (0 dB response represents 1 rad). The front excitation is only a little more effective than that at the rear in respect of the wobble-mode response.

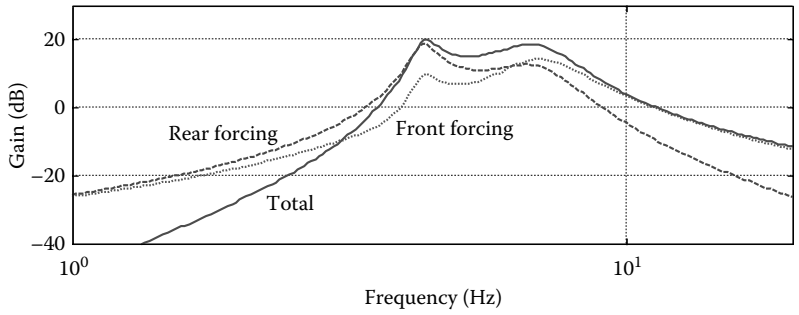


FIGURE 35.41 Steer angle frequency responses of standard bike at 57 m/s speed and 20° roll angle to road forcing in m (0 dB response represents 1 rad) with the standard steering damper coefficient doubled.

Figures 35.41 and 35.42 correspond to Figures 35.39 and 35.40, respectively, and they show the influences on the responses of doubling the steering damper coefficient. The change is effective in reducing the wobble-mode peak by about 12 dB, but it raises the weave-mode peak by about 4 dB. The steering damper is helpful to the wobble mode but damaging to the weave, as can also be seen in the root loci (Figure 35.22). The damper can be a rather critical component due to this conflict between the modal responses, and alternatives are of current research interest (see Section 35.5.9).

The possibility of nonlinear responses to road forcing being problematic arises naturally, since the motorcycle equations of motion are very complex and the terms of all orders are contained in them. Subharmonic, superharmonic, internal, and combination resonances [60] may occur. These nonlinear phenomena are potentially significant because they are likely to be outside the normal experience of riders, who

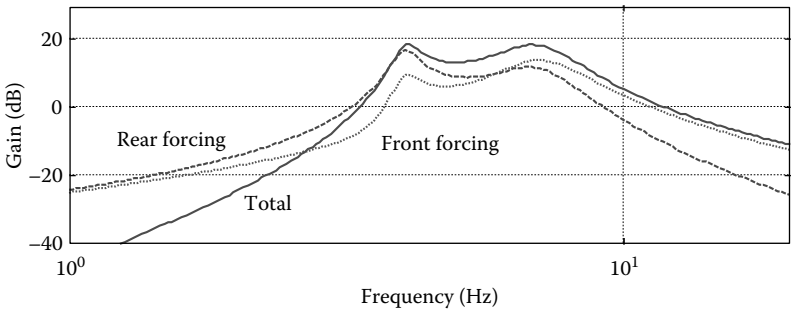


FIGURE 35.42 Steer angle frequency responses of standard bike at 47 m/s speed and 20° roll angle to road forcing in m (0 dB response represents 1 rad) with the standard steering damper coefficient doubled.

are unlikely to have learned how to deal with them, if they do occur. The phenomena have been studied by simulation and time-frequency analysis of computed motion time histories [61,62]. Large-amplitude subharmonic and superharmonic oscillations can occur, associated with forcing frequencies of twice or three times, or one-half or one-third of the weave or wobble natural frequencies, respectively. Superharmonic resonances were easier to provoke than subharmonics. Internal resonance requires whole-number relationships between natural frequencies and will not be notable unless the (linear) modal damping factors are low. Running conditions for which the wobble frequency is twice the weave frequency typically do occur, and simulations show that exciting one mode directly, by forcing at the natural frequency, can lead to a strong response in the other mode. However, the nonlinear energy transfer mechanisms appear relatively weak, and no real events can be positively associated with nonlinear behavior as yet.

35.5.8 Steering Compensation

It has been established both experimentally [15] and theoretically [22,50] (see Figures 35.22, 35.41, and 35.42) that a steering damper increases the damping of the wobble mode while decreasing that of the weave mode at high speeds. One possible way of improving on a standard steering damper is to make it speed adaptive [63], but variability according to speed does not directly address the fundamental problem. Stiff-framed machines are liable to have minimally damped weave and wobble modes at similar high speeds, as has been demonstrated earlier (Figures 35.18 and 35.19). The basic question is, "can we find an alternative to the damper, providing the same benefit to the wobble mode but not destabilizing the high-speed weave?" Searches for a passive mechanical compensator to improve on a standard damper are reported in [56,57] and for an active system for the same purpose, in [64]. In these studies, the steering damper is abstracted from the mechanical system and treated as an example of a compensator that responds to steering velocity by generating a torque between the main frame and the steering assembly (Figure 35.43). More general compensators made from springs, dampers, and inerters, the inerter consisting of a geared flywheel giving torque proportional to the relative angular acceleration across it [65], are considered by Nyquist frequency-response-based methods, followed by optimization over a wide range of operating conditions. Networks with good properties are identified, one of which is shown in Figure 35.44. This network operates like a steering damper at the wobble-mode frequency but behaves more like an inerter for the weave mode. Network parameters and frequency response properties and the influences on the root loci and road forcing responses can be seen in the references.

In the case of the active compensator, passivity is no longer a restriction on the achievable performance, the compensator may be of high order without becoming impractical, and adaptation of the parameters to operating conditions is quite feasible. The compensator design in [64] is achieved by H-infinity loop-shaping procedures [66–68], and further advantage with respect to the standard damper and the passive network is demonstrated. Work in progress is directed toward understanding of the influences of acceleration and deceleration. Also, practical evaluation of passive and active systems in the laboratory and on running motorcycles is in progress.

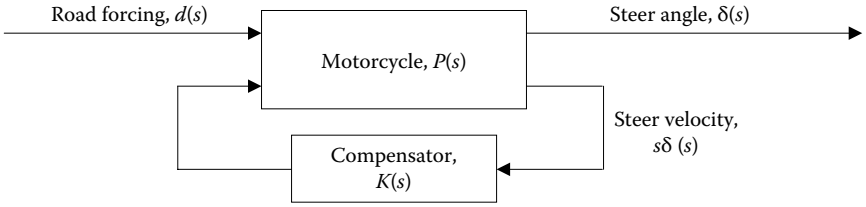


FIGURE 35.43 Abstraction of steering compensator from the motorcycle for generalized treatment. (From Evangelou, S. et al., Steering compensation for high-performance motorcycles, *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, pp. 749–754, December 14–17, 2004; Evangelou, S. et al., *ASME J. Appl. Mech*., in review.)

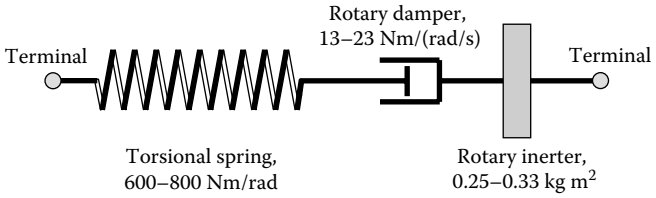


FIGURE 35.44 Passive compensator with advantageous properties. (See Evangelou, S. et al., Steering compensation for high-performance motorcycles, *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, pp. 749–754, December 14–17, 2004; Evangelou, S. et al., *ASME J. Appl. Mech*., in review for determination of parameter values.)

35.5.9 Acceleration and Deceleration

Many motorcycles are capable of high levels of acceleration and deceleration. Often, stoppies set the limit on what is possible in braking, while wheelies or engine power set the acceleration limits. From a steady condition, application of driving or braking brings an in-plane response directly [48], and the lateral dynamics are affected, partly by the redistribution of tire loadings and partly by virtue of the inertial influences newly introduced. Strictly, the motions cannot be analyzed by considering small-perturbation-from-trim states, since trim states require the speed to be constant. However, with appropriate reservations, frozen-time eigenvalue analysis can be used to predict behavior [36]. Acceleration and deceleration are not just interesting from an analytical standpoint; they are important practically, due to safety considerations.

First, let us examine the transient in-plane behavior when accelerating suddenly from low speed [48]. Figure 35.45 shows the standard machine traveling straight at trim at 5 m/s when a demand for 5 m/s/s acceleration occurs. The speed controller responds by applying a driving torque to the transmission sprocket, and the motorcycle attitude adjusts to the new situation. The rear tire longitudinal slip settles to just under 0.03. The main frame bounce is positive downward, and positive pitching gives a nose-up attitude. The swing-arm angle is positive in rebound. The corresponding wheel loads are shown in Figure 35.46.

Severe braking by the front wheel only from 75 m/s straight-running trim, with a deceleration demand of 10 m/s/s, is shown in Figures 35.47 and 35.48. In decelerating, the aerodynamic force system helps, but nevertheless, in this case, the front tire shear force needs to be greater than the front wheel load to achieve the deceleration objective. The front tire slip ratio reaches below −0.05, the tire getting quite near to its adhesion limit. The rear tire continues to contact the ground, the load reducing to only 200 N at one stage.

If the motorcycle brakes were applied sufficiently to cause wheel locking, the major influence on the lateral motion of the motorcycle would come from the completely changed behavior of the tires, when they stop rolling. It is easy to show, by representing the tire lateral forces for locked wheels as only marginally effective, that the vehicle will become rapidly divergent, if either front or rear wheel locks. In the former case, the rider's control mechanism is effectively removed by the wheel

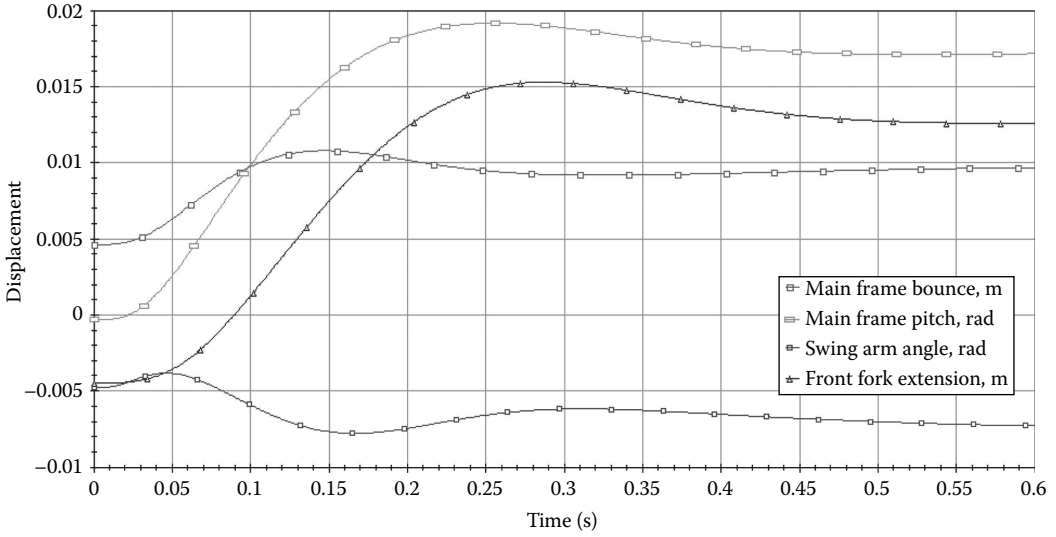


FIGURE 35.45 Motorcycle responses to step input of 5 m/s/s acceleration demand from 5 m/s straight-running trim state.

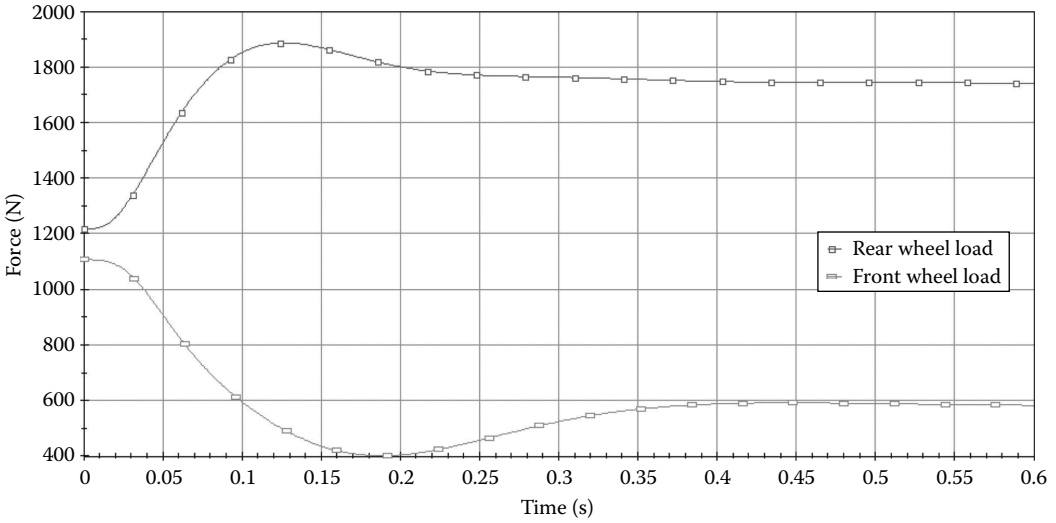


FIGURE 35.46 Rear and front wheel loads for the simulation run of Figure 35.45, showing the transfer of load from front to rear corresponding to the acceleration of 5 m/s/s. The manner in which suspension kinematics influence the transient load transfer is illustrated in (From Sharp, R.S. et al., *Advances in Computational Multibody Systems*, J. Ambrósio ed., Springer-Verlag, Dordrecht, the Netherlands, pp. 45–68, 2005).

locking. The only possible salvation lies in releasing the brake and getting the front wheel to roll again. In the latter case, control is possible through the use of the steering system to influence the front tire, but the stabilizing task is likely to be difficult. Effective antilocking systems have a particular appeal in the world of single-track vehicles, since locking wheels is so problematic.

Now, let us examine the influences of acceleration and braking on the root loci of the straight-running motorcycle through the speed range. As described earlier, the calculations can be done in either of two ways. The first way involves performing a simulation run in which the machine accelerates from low speed at a constant rate, avoiding a wheelie, to a speed determined by the engine power available, and then carrying through a root locus computation, as if the simulation

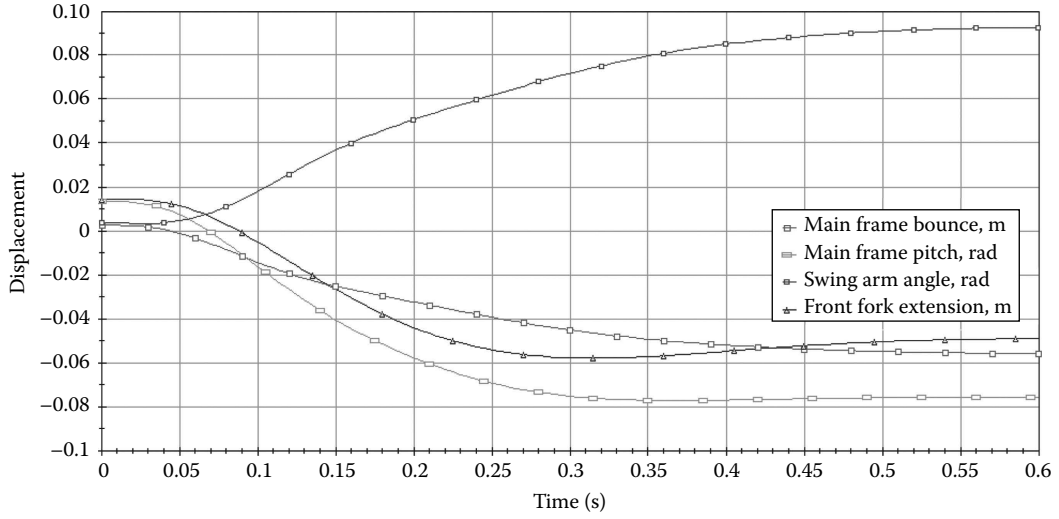


FIGURE 35.47 Motorcycle responses to step input of 10 m/s/s deceleration demand from 75 m/s straight-running trim state by front braking only.

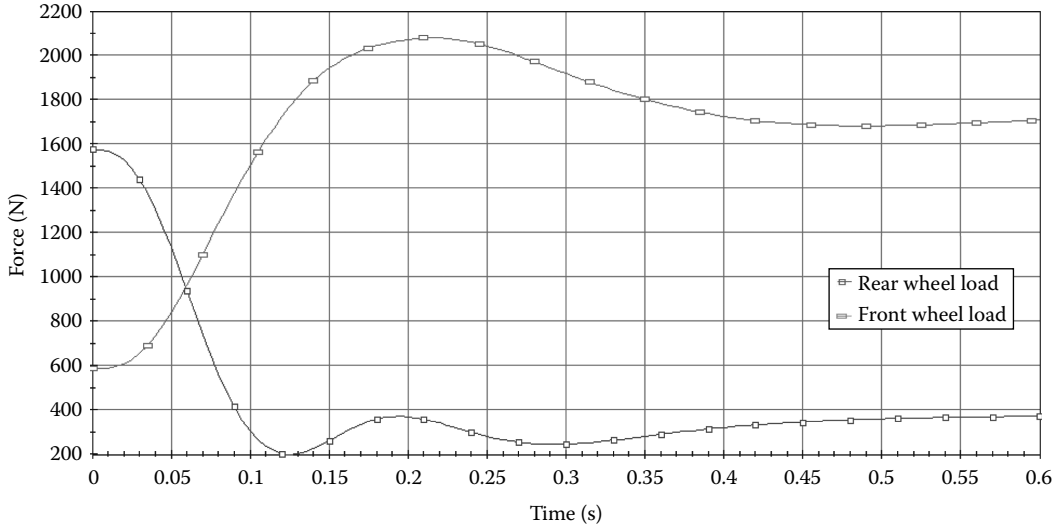


FIGURE 35.48 Rear and front wheel loads for the simulation run of Figure 35.47, showing the transfer of load from rear to front corresponding to the deceleration of 10 m/s/s. The front tire shear force is greater than the wheel load in this run. (See Sharp, R.S. et al., *Advances in Computational Multibody Systems*, J. Ambrósio ed., Springer-Verlag, Dordrecht, the Netherlands, pp. 45–68, 2005 for suspension kinematic design influences on the load transfer.)

run were a succession of trim states. This is the "frozen-time" method [36]. The second way involves the application of an artificial horizontal gravitational field, so that each mass in the system is acted on by a corresponding force opposing the (imagined) acceleration, with the simulation run to establish trim states employing only a marginal acceleration, typically 0.1 m/s/s. The purpose of the speed change is to cover the range of speeds of interest in a single simulation run, with negligible acceleration influences. Fortunately, the differences between the two methods do not have to be studied too deeply, since the root loci resulting are indistinguishable. Figure 35.49 contains frozen-time results for accelerations of 2.5 and 5 m/s/s, while Figure 35.50 shows similar results for decelerations of 5 and 10 m/s/s. Bearing in mind the different speed ranges covered,

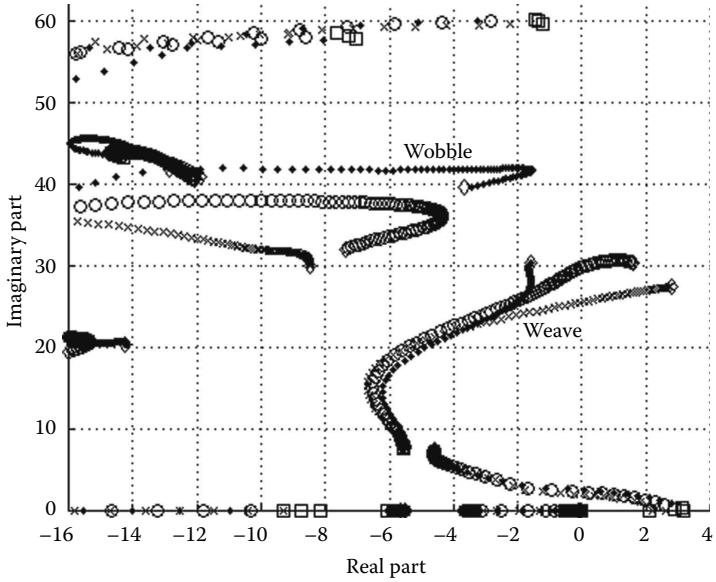


FIGURE 35.49 Root loci for straight running at constant speed from 0.5 to 75 m/s (dots), with acceleration of 2.5 m/s/s from 0.4 to 62 m/s (circles) and with acceleration of 5 m/s/s from 0.4 to 50 m/s (crosses). The speed ranges are limited by the engine power output of 88 kW. A square represents the low-speed end of each locus, while a diamond represents the high-speed end.

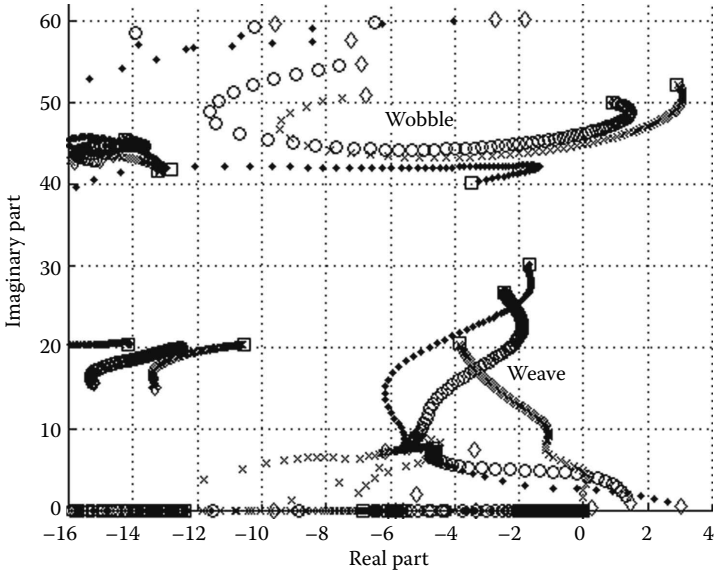


FIGURE 35.50 Root loci for standard machine in straight running at (almost) constant speed from 75 to 0.3 m/s (dots), with deceleration of 5 m/s/s (circles) and with deceleration of 10 m/s/s (crosses). A square represents the high-speed end of each locus, while a diamond represents the low-speed end.

the acceleration can be seen to destabilize the high-speed weave and to stabilize the wobble mode and conversely. The influences of acceleration and deceleration on the stability of motorcycles in cornering have yet to be studied in any depth.

If the machine is in a trim state at 20 m/s speed and 20° roll angle, a sustained steer torque of −4.9 N m is necessary to maintain the trim. The trim state is stable, so that removing the steer torque input can be expected to result in the machine coming upright and straightening course.

To show the influence of accelerating or decelerating on such a turn exit, the simulation model is set up to accelerate at 1 m/s/s by driving the rear wheel and to maintain speed or to decelerate at 1 m/s/s by braking the front wheel (Figures 35.51 and 35.52).

Accelerating at the same time as removing the trim steering torque causes the steering system to return to the zero position more quickly. The influence on the machine's roll angle is to slow down the return to straight running. It is a common experience that a motorcycle feels good when accelerating out of a turn, and these results suggest that the slowing down of the rolling-to-upright motion is the basis for the attractive feel. Skillful riders will apparently coordinate throttle and brake to obtain a smooth transition from a turn to a straight path.

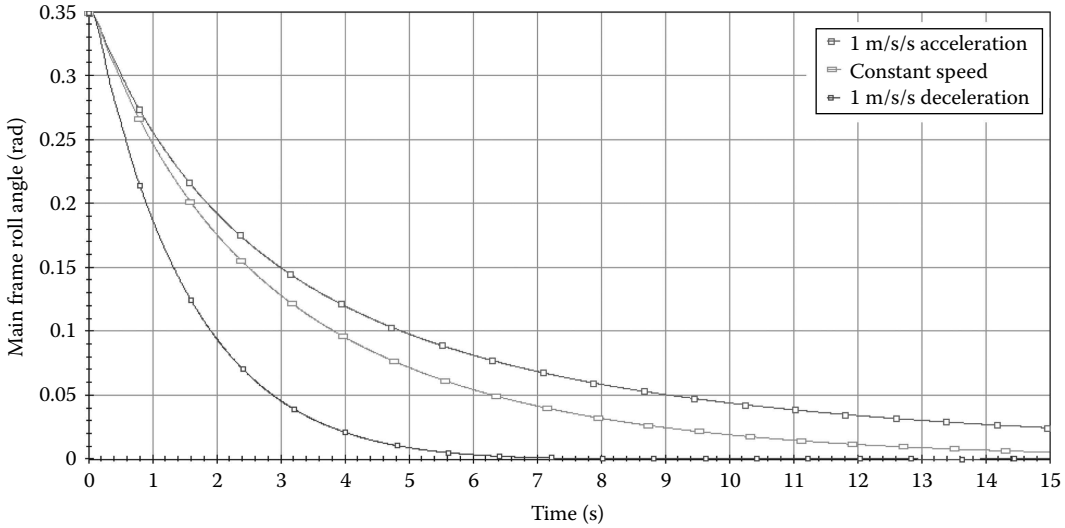


FIGURE 35.51 Motorcycle roll angle histories with 1 m/s/s acceleration, constant speed, and 1 m/s/s deceleration from initial 20° roll angle and 20 m/s speed, with trim steer torque removed.

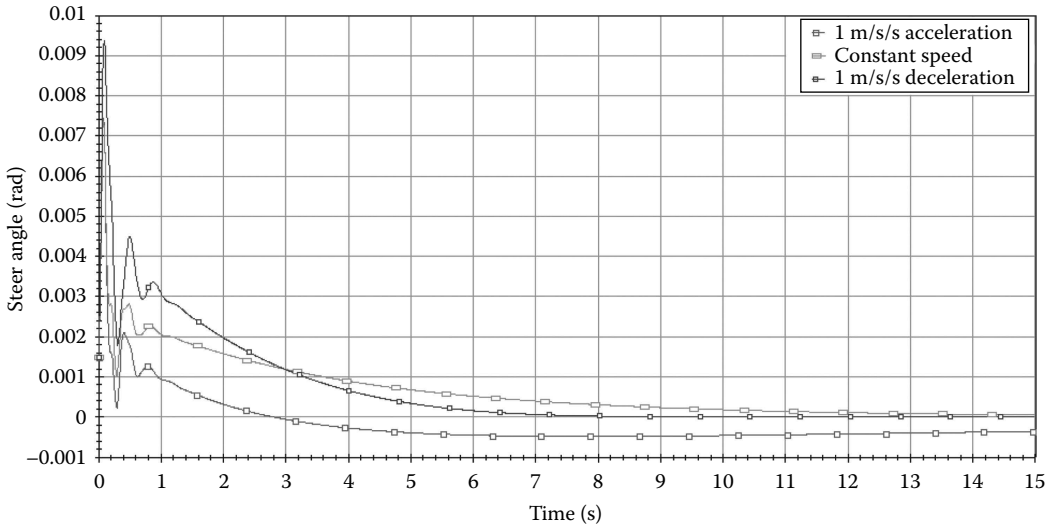


FIGURE 35.52 Motorcycle steer angle histories with 1 m/s/s acceleration, constant speed, and 1 m/s/s deceleration from initial 20° roll angle and 20 m/s speed, with trim steer torque removed.

35.5.10 Power Accounting

The power balance check calculations described in Section 35.4.13 allow the determination of all the sources of power supply and consumption in the system [47,50]. Aerodynamic losses are relatively simple, but the losses due to various aspects of the tire force and moment systems are not so obvious. Results from the standard motorcycle for a sustained 45° roll angle turn with a slow acceleration from 10 to 70 m/s, achieved in about 200 s, are shown in Figures 35.53 and 35.54. Power supplied to the drive sprocket by the engine is lessened a little by the rear tire losses before contributing to the vehicle motion. Aerodynamic losses account for a high proportion of the total power consumption. At 70 m/s, 92% of the engine power goes to propelling the motorcycle and 79% of the total is dissipated aerodynamically. The front tire rolls freely and there are hardly any losses associated with the longitudinal and lateral forces. However, the

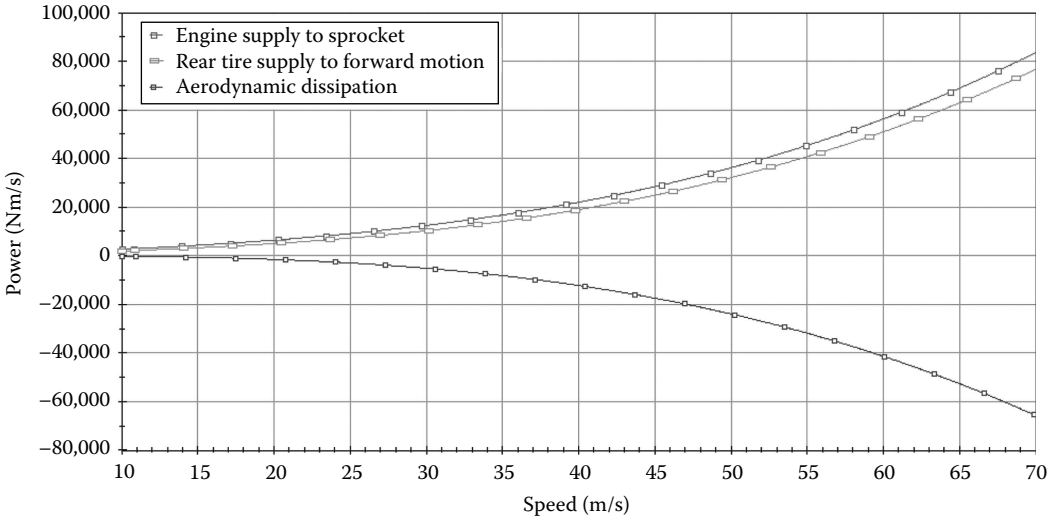


FIGURE 35.53 Power supply from engine and from rear tire longitudinal force and power dissipation by aerodynamic influences in sustained turning at 45° roll angle with 0.3 m/s/s acceleration, as functions of speed.

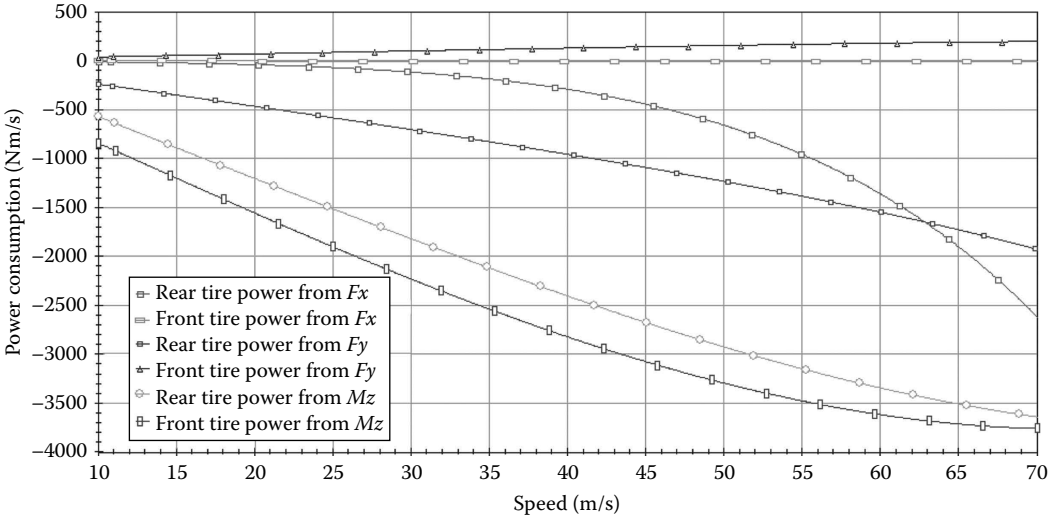


FIGURE 35.54 Tire power consumption components in sustained turning at 45° roll angle with 0.3 m/s/s acceleration, as functions of speed.

power consumption due to the front tire aligning moment builds up to about 3.8 kW at 70 m/s. The rear tire power consumption shows a typical strong dependency on longitudinal force, since the power due to *Fx* increases with force, slip, and speed, and the slip increases as the force does.

35.6 SENSITIVITIES TO CONFIGURATION PERTURBATIONS

35.6.1 Experimental Configuration Changes

Bayer [15] carried out systematic experiments to determine weave-mode damping factors for 13 different motorcycles through their speed ranges. His results are shown in Figure 35.55, in which the variations observed are indicated by the hatching lines. Three machines had only modest speed capability, which machines are designated "low-speed machines" in the figure.

Bayer also prepared a special high-performance motorcycle for testing, with the possibility to change the wheelbase by lengthening the swing arm and to vary the steering head angle and the trail. A potential problem in such testing is that unknown variations in stiffnesses will occur when the geometry is modified and these may be sufficient to spoil the interpretation of the results, when these influences are considered negligible. Results for 40 and 46 m/s show that increasing the wheelbase and using a larger steering head angle (more rake) are both beneficial. At the higher speed, increasing the trail also raised the damping factor but, at the lower speed, this change made no difference. At both test speeds, raising the steering system inertia was quite strongly destabilizing. Bayer produced a bar chart showing the main influences on the weave damping, which chart is reconstructed as Figure 35.56 (see [22] also). The key to the modifications made to the datum machine (at 36 m/s) is shown in Table 35.1. Bayer also found the wobble-mode frequency to lie between 9 and 9.5 Hz and the weave-mode frequency to rise from 2 to 4 Hz as the speed increased from 26 to 54 m/s.

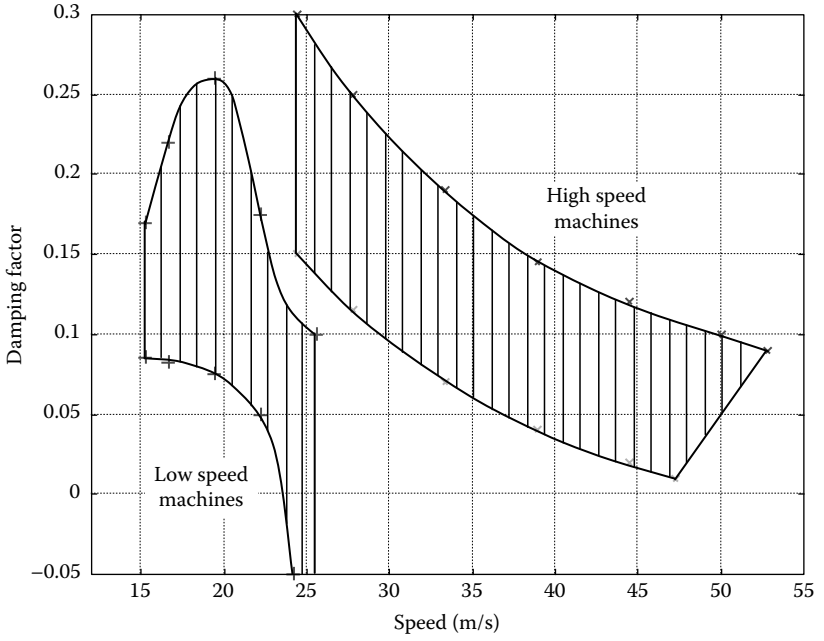


FIGURE 35.55 Weave-mode damping factors determined by experiment for 13 different motorcycles by Bayer (From Bayer, B., *Automobil_Industrie*, 2, 193, 1988). The shading indicates the spread of results obtained.

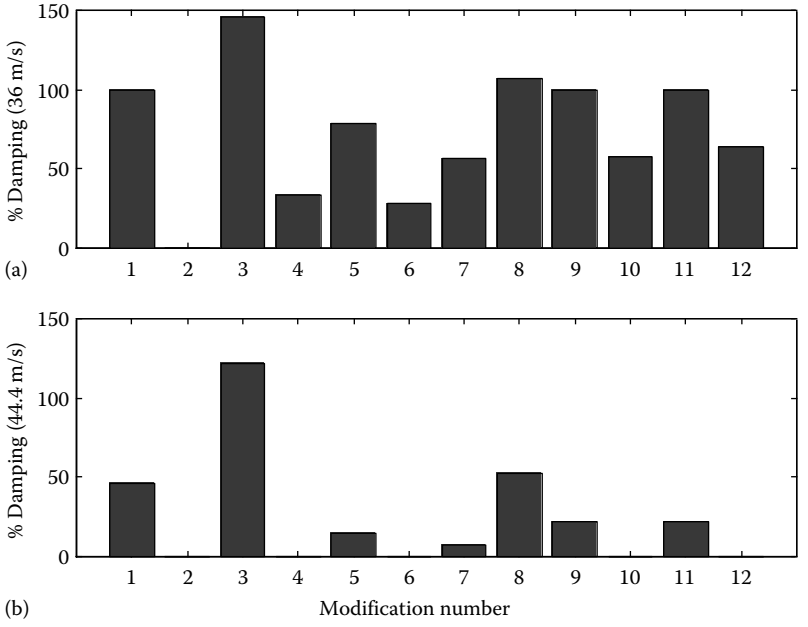


FIGURE 35.56 Damping of weave mode relative to datum configuration (100%) as found experimentally by Bayer (From Bayer, B., *Automobil_Industrie*, 2, 193, 1988) for 36 (a) and 44.4 m/s (b) speed, with modifications to test motorcycle as defined in Table 35.1. Zero entries indicate instability.

TABLE 35.1
Motorcycle Modifications by Bayer

1	Datum case
2	Rider's feet located on the pillion passenger's footrests
3	With the addition of a pillion passenger
4	10 kg mass added to each of two saddlebags
5	Top box, containing 5 kg mass, added
6	5 kg mass added to the front forks
7	Front fairing added
8	Front fork stiffener added
9	Steering damper added
10	Steering head bearings overtightened
11	Front wheel unbalanced
12	Improper tires employed

Source: Bayer, B., *Automobil_Industrie*, 2, 193, 1988.

35.6.2 Theoretical Configuration Changes

Some parameter change influences on the damping factors of the weave and wobble modes for the motorcycle and rider at 60 m/s are shown in Figures 35.57 and 35.58, respectively. In Figure 35.59, corresponding variations in the capsize mode eigenvalue are given. The nominal motorcycle and rider system is the first case treated, after which one parameter is raised in turn by 10% of its nominal value. The changed parameter is depicted in short form on each chart and listed more fully in Table 35.2.

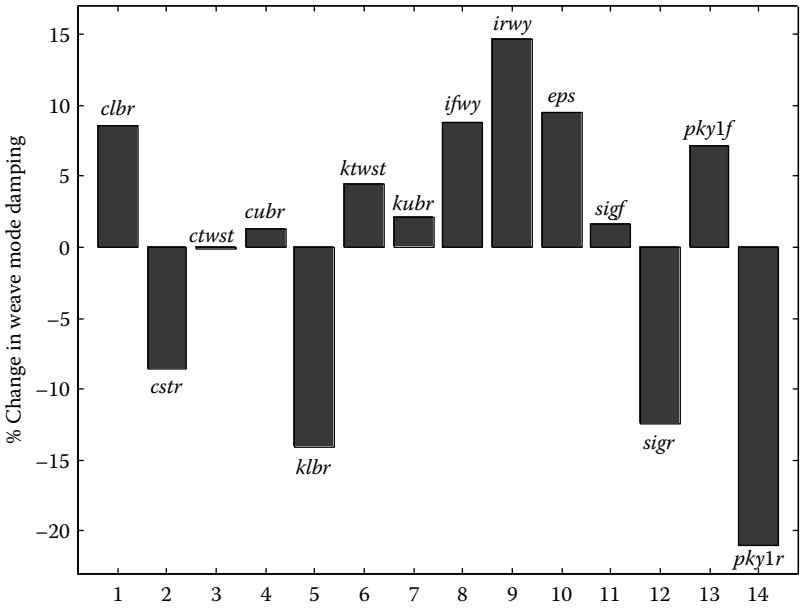


FIGURE 35.57 Bar chart showing how the weave-mode damping factor for 60 m/s speed in straight running is affected by 10% increases in individual parameter values. Variants are defined more fully in Table 35.2. The weave-mode damping factor at 60 m/s is 0.0563.

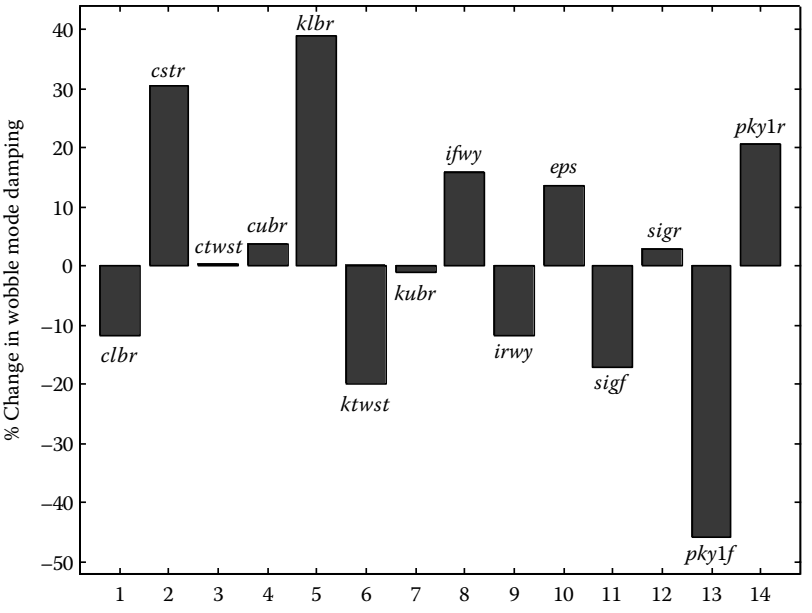


FIGURE 35.58 Bar chart showing how the wobble-mode damping factor for 60 m/s speed in straight running is affected by 10% increases in individual parameter values. Variants are defined more fully in Table 35.2. The wobble-mode damping factor at 60 m/s is 0.0384.

The weave-mode damping is quite sensitive to the rider lower body stiffness and damping, in opposing senses—more damping and less stiffness are helpful. Rider to rider variations in system behavior can be expected based on this result. The steering damper is again shown as deleterious to the weave mode. Increased spin inertia in each road wheel helps, and it follows that there is some advantage, in this context, from the convention that the engine and transmission rotors spin in sympathy

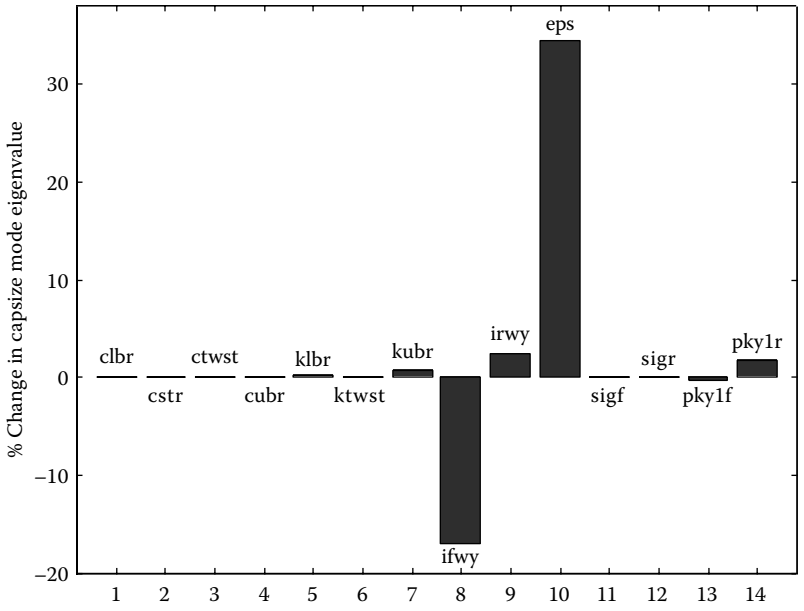


FIGURE 35.59 Bar chart showing how the capsize mode eigenvalue for 60 m/s speed in straight running is affected by 10% increases in individual parameter values. Variants are defined more fully in Table 35.2. The nominal machine has eigenvalue −0.041 at 60 m/s.

TABLE 35.2
Parameter Variations in Figures 35.57 through 35.59

Model	Parameter Increased by 10% of Nominal Value
1	Rider lower body damping coefficient
2	Steering damper coefficient
3	Frame torsional damping coefficient
4	Rider upper body damping coefficient
5	Rider lower body stiffness
6	Frame torsional stiffness
7	Rider upper body stiffness
8	Front wheel spin inertia
9	Rear wheel spin inertia
10	Steering head angle
11	Front tire relaxation length
12	Rear tire relaxation length
13	Front tire cornering stiffness
14	Rear tire cornering stiffness

In each case, the nominated parameter is increased by 10% of its nominal value.

with the rear wheel rotation. An increased steering head angle is helpful. The rear tire is much more influential than the front tire, a short relaxation length and, surprisingly, a low cornering stiffness at the rear being good. The front tire cornering stiffness should be high. These results offer some insight into the weave-mode damping changes that often occur when the tires on a motorcycle are exchanged.

The wobble-mode damping is improved by increases in steer damping, rider lower body stiffness, front wheel spin inertia, steering head angle, and rear tire cornering stiffness. The damping factor is reduced by increases in rider lower body damping coefficient, frame torsional stiffness,

rear wheel spin inertia, front tire relaxation length, and front tire cornering stiffness. The last is the most influential parameter, and again, the significance of the rider lower body parameters is notable. These rider parameters act oppositely on the two oscillatory modes, as do the steering damper and the various tire parameters. An increase in the steering head angle benefits both oscillations, but it has a negative influence on the capsize mode. Note that the steering head angle is defined such that the location of key points is not affected when it is altered, which implies that the trail increases as the head angle does. The other parameter affecting the capsize significantly is the front wheel spin inertia, which should be large for stability. However, at high speeds, the steering torque needed in a transient maneuver is dominated by the gyroscopic influence of the front wheel [58], so that there would be a price to be paid in heavy transient steering at high speeds from having a high-spin-inertia front wheel. Other such sensitivity results can be found in [22].

The only direct comparison possible with Bayer's results is in respect of the steering damper addition and even this cannot be quantified, since the coefficient of Bayer's damper is not known. However, it is the case that Bayer's steering damper did reduce the weave-mode damping factor, as the theoretical results regularly show, but it only did so by a little.

35.7 OPTIMAL LINEAR PREVIEW RIDER CONTROL AND HANDLING QUALITIES

35.7.1 Motorcycle Rider Model Background

The purposes in modeling the motorcycle rider are to enable path-based simulations and to understand rider control and the matching of rider to machine better. The problem of reproducing in the virtual world the skill of a racing rider, capable of exploiting the capabilities of a high-performance motorcycle fully, is challenging. Such a virtual rider would enable the discovery of the maneuvering limits of alternative designs of systems and act as an aid to optimal design of products. The applicability to racing and product improvement generally is clear. The usefulness of a rider model depends on its computational complexity and speed of calculation and its robustness against variations in running conditions and machine design. The most practical rider model is likely to have to compromise between performance, speed, and robustness.

In the background, we know about the following features of the motorcycle riding problem: (a) The motorcycle is divergently unstable at very low speeds. The very-low-speed instability becomes oscillatory before it stabilizes, which will typically happen at about 6 m/s. After that, the system has its capsize mode near to the stability boundary. To a human rider, in this higher speed range, it does not matter much whether or not the machine is self-stabilizing, since it can be stabilized by the person without much effort in any case. (b) The motorcycle has a nonminimum phase roll response. In response to a step input of steering torque to the right, it will reach a left-hand-turn steady state, after a transient phase. (c) A motorcycle rider needs a particularly large amount of road preview [69]. If a driver of any road vehicle has too restricted a view of the roadway forward, driving is not feasible, even though the lateral position of the vehicle relative to the roadway may be quite clear. Restricted preview of the road, by fog or poor headlights, for example, makes accurate path following at reasonable speeds difficult and energy consuming. Path following by single-point preview, that is, with knowledge of the lateral location of the path at only one point ahead of the vehicle, is critically dependent on the preview distance. If the preview distance is short, tracking will be tight but oscillatory, while if it is long, tracking will be loose or sluggish, with much corner cutting. The conflict between short and long previews is avoidable by using multipoint preview, with the rider knowing effectively the whole of the relevant path at any point of a journey. How much of the path ahead is relevant depends on the vehicle dynamics. The faster are the plant dynamics, the lower will be the required preview time. (d) Steering control acts largely to yaw the relatively large longitudinal velocity component and to thereby influence the vehicle lateral position significantly only some way ahead of the current location. Due to this actuation mode, the future path requirements are a

much more useful information to the rider than are the current ones. (e) A motorcycle rider has two steering control inputs, steering torque (or angle) and rider lean torque (or angle), although there is much evidence to suggest that rider leaning is a very weak control in comparison with steering torque [8]. (f) A motorcycle rider, in common with other drivers, can choose the level of concentration to be applied to and accuracy associated with the task. Tight control is characterized by high effort and accuracy and loose control with the converse. Only in racing is tight control demanded almost the whole time. (g) The virtual motorcycle can be stabilized over a very full range of operating conditions by PID feedback of roll angle error to steer torque (see Section 35.4.10).

35.7.2 Existing Motorcycle Rider Models

The earliest effective work on modeling the motorcycle rider was by Weir, who viewed the problem as one of regulating against disturbances that could not be anticipated and following a path that could only be known at the current longitudinal location of the motorcycle [70,71]. The rider was presumed to act on some motion information and to feed it back to steer torque control to achieve the required regulation. Prior research on manual control in regulatory tasks was applied to the problem. The main results used were that the operator adopts a transfer characteristic reflecting physical attributes like information processing, decision delays, and neuromuscular delays and adjusts parameters so that the man–machine open-loop frequency response function for each closed loop is of the form

$$G_r(\omega)G_m(\omega) = \frac{\omega_c e^{-j\omega\tau}}{j\omega} \tag{35.42}$$

in the region of the gain "crossover" frequency. Here, $G_r(\omega)$ is the rider response, $G_m(\omega)$ is the machine response, ω_c is the crossover frequency, and τ is the time delay. For each loop, the rider supposedly increases his or her gain sufficiently to give the closed-loop system the desired performance, thus setting the crossover frequency. Note that when $\omega = \omega_c$, the gain is unity. Eaton's experiments with a 175 cc motorcycle, rider braced to prevent leaning, moderate speeds only, and impulsive excitation [2] allowed the identification of parameter values for a roll angle to steer torque feedback that accorded closely with the crossover theory, lending support to it. Weir also explored additional loop closures to provide heading and lateral position controls, using rider lean angle as actuation. Using this background, Lenkeit constructed a servo-rider [72], choosing roll angle over steering angle control for low speeds and roll angle to steering torque control for higher ones.

Most other motorcycle rider models are aimed more directly at path following and recognize the need for preview of the path. In [73,74], control torques are applied on rider lower body, rider upper body, and steering system, each a linear combination of roll angle and a path preview parameter. This parameter was a weighted integral over all the path ahead, containing a preview distance to be chosen conveniently. This model also was shown capable of matching experimental results. In the later paper, steering torque proportional to roll velocity was added to the control scheme, to improve low-speed operation. A brief description of a somewhat similar steering torque controller with proportional and derivative terms deriving from roll angle and terms proportional to path error and path slope, relative to the motorcycle heading, at a single preview point ahead can be found in [20].

Steer torque with terms proportional to steer velocity, roll angle error, and roll velocity was used in [75]. Single-point preview was converted into a roll angle demand, and effective gain values were found by surveying operation over speed and roll angle ranges. A fuzzy controller was devised to smoothly connect or schedule the gains, according to the operating conditions. It was found necessary to add a new term with a somewhat obscure physical basis, called a "hold" torque, to get good performance.

Notable among motorcycle rider models are those of Frezza et al. [76]. For the design of the controls, which include steer angle and longitudinal tire forces, the motorcycle is represented as having two rigid frames, joined at a vertical steering head and having pure-rolling tires. Such a machine

has its path over the ground determined by its steer angle and speed histories only. A differential equation relates the roll angle, the speed, and the path curvature. This roll angle, speed, and path curvature relation is used to constrain the solution of a nonlinear optimization problem, in which a weighted quadratic function of path error and roll angle error is minimized over a chosen preview time, subject to upper bounds on magnitudes of steer angle and roll angle. The desired roll angle is the steady-state solution of the roll, speed, and path curvature equation, with the desired values of speed and curvature being used as input to the equation. The roll error term serves to couple the longitudinal and the lateral dynamics in an interesting way. The local path immediately in front of the motorcycle is represented in a discrete form through the parameters of B splines. Therefore, the optimal solution gives these parameters and the roll angle history. The required steer angle control is derived from the optimal solution, and it is applied for a single time step, after which all the computations are repeated. The strategy can be classed as a model predictive control one, with faster computation than would be obtained by calculating with the full motorcycle model throughout. Important limitations on achievable performance associated with tire force saturation are necessarily represented as constraints, some of which appear to be decided arbitrarily. Also, applying the final steer angle and longitudinal force controls to a complex multibody motorcycle model involves faith and hope that they will function well. The authors testify to good tracking results in virtual trials but do not address speed of computation issues. Extending the preview horizon is likely to improve performance but degrade computation speed. Much compromise appears to be involved.

Another commercial rider model is that included in BikeSim [49], which uses a development of the detailed multibody motorcycle model described in [47,48]. The rider model capitalizes on the fact that the motorcycle is quite easy to stabilize over a wide operating range with PID roll angle to steer torque feedback with speed adaptive gains. With this notional control scheme in place, the path-tracking problem becomes similar to that for a car or a truck. For this class of problem, MacAdam's strategy [77] is long developed and tested. According to this strategy, a simplified linear model of the vehicle is contained within it. This linear representation is used at a particular time step to solve the optimization problem, "What is the value of the fixed steer angle that minimizes the path-tracking error over the preview horizon?" In solving this problem, account is taken of the motion state at the instant and the path in front of the vehicle, relative to the vehicle, that is, the local path, over the preview distance. As with Frezza's method, the steer angle is held for a single time step and the computations are repeated. Again, as with Frezza's method, the optimizing calculations have to be done online and they are simplified in the interests of rapidity. Again, Virtual trials are reportedly yielding good results [78].

Although the rider models described have considerable capabilities, it is thought that they are not the ultimate solution to the problem either of computing the best possible control or that of computing the fastest reasonably effective control. It is considered that optimal preview control offers something both better and faster. Only linear theory has so far been developed [79,80], but bearing in mind that the series expansion linearization method can be applied to the motorcycle operating in the neighborhood of any trim state, linear controllers that are adaptive to the operating conditions appear potentially powerful. Let us consider a foundation for linear quadratic optimal preview control theory in Section 35.7.3 as follows. In what follows, we assume constant speed and two control inputs, steering torque and rider upper body roll torque.

35.7.3 Outline of the Application of Optimal Linear Preview Regulator Theory

The view of the roadway and the vehicle motion taken is illustrated in Figure 35.60. The track centerline is presumed known relative to the fixed ground, and the vehicle motion is similarly considered in a fixed or absolute reference system. The relationship between the track and the vehicle is known, since both are known absolutely. The motorcycle is represented by its constant-coefficient linear equations of motion for small perturbations from a trim state, in this case involving straight running at constant speed. The rider exercises control by applying steer torque to the handlebars, reacted on his or her upper body, and upper body roll torque, reacted on his or her lower body.

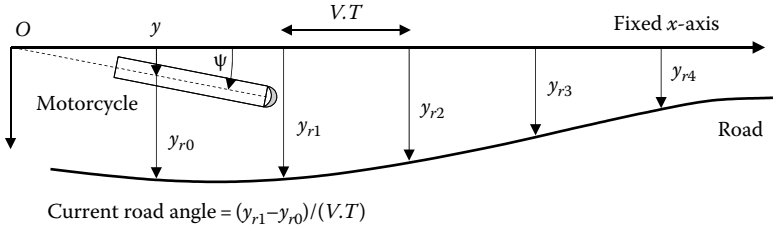


FIGURE 35.60 Representation of a motorcycle traveling along a defined path in absolute terms, with speed V and discrete time step T.

Initially, the road is presumed to follow the line of the x-axis in a general sense, so that there will be only small differences between the attitude angle of the motorcycle and that of the road. The motion is viewed in discrete time, to deal advantageously with the time delays evident in the system.

With the transition of the system through one time step, the motorcycle moves along the road by distance VT. Relatively speaking, the road moves toward the motorcycle by that distance unit. The discrete-time road dynamics are those of a shift register. The combined system of vehicle and road can be described by joining together the motorcycle dynamics and the road dynamics, which two subsystems become coupled through the path-tracking objective function, which expresses the idea that costs will be attached to differences between y, the lateral position of the motorcycle reference point, and y_{r0}, the corresponding lateral position of the track (see Figure 35.60), at each time point in a journey. Other similar measures of tracking error may also be included in the objective function, and a cost must be associated with control power, with weighting parameters to specify relative importance.

The first applications of this form of optimal linear preview control theory were to car steering control [81,82]. More background can be found in those references. An outline of the theory follows.

Suppose our linear motorcycle model of any degree of complexity has the discrete-time form

$$\begin{aligned} x_v(k+1) &= A_v x_v(k) + B_v T(k) \\ y_v(k+1) &= C_v x_v(k) + D_v T(k) \end{aligned} \tag{35.43}$$

Imagine that the lateral profile of the road is defined by discrete sample values, with values from past observations of the road ahead being stored as states of the full vehicle/road system. As the system moves forward in time, a new road sample value is read in and the oldest stored value is discarded, corresponding to the vehicle having passed the point on the road to which this oldest value refers. All the other road sample values are shifted through the time step, nearer to the vehicle. The dynamics of this shift register process are represented mathematically by

$$y_r(k+1) = A_r y_r(k) + B_r y_{ri}(k) \tag{35.44}$$

$$\text{where } A_r \text{ is of the form : } \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{and} \quad B_r \text{ is of the form : } \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 1 \end{bmatrix}$$

Combining vehicle and road equations together, we obtain the full dynamic system description:

$$\begin{bmatrix} x_v(k+1) \\ y_r(k+1) \end{bmatrix} = \begin{bmatrix} A_v & 0 \\ 0 & A_r \end{bmatrix} \begin{bmatrix} x_v(k) \\ y_r(k) \end{bmatrix} + \begin{bmatrix} B_v \\ 0 \end{bmatrix} T(k) + \begin{bmatrix} 0 \\ B_r \end{bmatrix} y_{ri}(k) \tag{35.45}$$

The complete problem is now in a standard discrete-time form

$$z(k+1) = Az(k) + Bu(k) + Ey_{ri}(k)$$

$$y(k) = Cz(k) + Du(k)$$

If y_{ri} is a sample from a white-noise random sequence, the time-invariant optimal control minimizing a cost function J, given that the pair (A, B) is stabilizable and the pair (A, C) is detectable [84], is

$$u^*(k) = -Kz(k), \quad \text{where} \quad K = \{R + B^T PB\}^{-1} B^T PA \tag{35.46}$$

where $J = \lim_{n\to\infty} \sum_{k=0}^{n} \{z^T(k)Qz(k) + T^T(k)RT(k)\}$ and P satisfies the matrix-difference Riccati equation

$$P = A^T PA - A^T PB\{R + B^T PB\}^{-1} B^T PA + C^T QC \tag{35.47}$$

in which Q is a diagonal weighting matrix, with terms corresponding to the number of performance aspects contributing to the cost function, and $R = \text{diag}[r_1, r_2]$, corresponding to the two control inputs, steering torque and rider upper body lean torque. Here, only one performance measure is used, $Q = q_1$ being set to represent the importance of tracking errors. r_1 and r_2 reflect the costs associated with steering torque and rider lean torque, respectively. Optimization calculations find compromises between high tracking accuracy and low utilization of steer torque and lean torque, according to the weights in the cost function.

From [82], it is also known that the optimal time-invariant control given by (35.46) is also optimal if the white-noise sequence representing the road lateral profile is low-pass filtered, to make it more realistic, and a "full" preview of the road [81,82] is available. It is implied that the white-noise disturbance assumption is not as restrictive as it might appear to be.

As in [81], the optimal control is found by partitioning the problem into non-preview and preview parts. Firstly, the non-preview, standard linear quadratic regulator (LQR) optimal control is found. Then, the state-transition matrix of the closed-loop system is applied recursively to the preview error samples to find the preview gains. The nature of the resulting optimal control is shown in Figure 35.61. There is a full-state-feedback part and a preview part, which superpose to give the optimal controls.

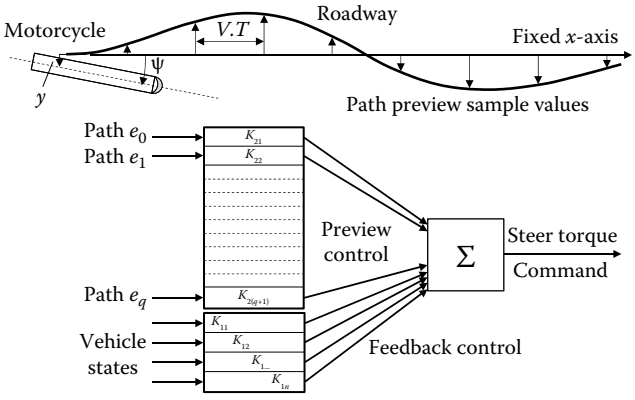


FIGURE 35.61 Illustration of the form of the optimal controls resulting from assuming a white-noise disturbance and an infinite optimization horizon. Standard LQR feedback parts and preview parts combine linearly to give the total controls.

35.7.4 Optimal Control Results

Several parameters are needed to define a particular control optimization problem. The machine speed must be specified, and the balancing between the different parts of the cost function representing use of steer torque, given by r_1; use of rider upper body lean torque, given by r_2; and tracking error, given by q_1, must be set. In addition, the number of preview points to be used needs to be decided. Many different results can be generated. Some representative cases follow. State-feedback gains are shown first in bar chart form, followed by graphs showing preview gains as continuous curves, although they are actually discrete values at the discrete-time interval of 0.02 s apart.

In Figures 35.62 and 35.63, gains are shown for speeds of 10, 40, and 70 m/s with 500 preview points, equal weights on the two controls, and a high cost on tracking errors, leading to tight control. In each case, the linearization relates to small perturbations from straight running and uses the appropriate trim state determined by simulation of the nonlinear system. As should be expected, the only nonzero state-feedback gains are associated with out-of-plane state variables. Symmetry of the straight-running condition implies zero gains for all in-plane states. Feedback gains relating to steer torque are much larger than those relating to body torque, the largest contributors, in both cases, being roll angle and frame twist angle (see Table 35.3 for a listing of the involved variables). The theory as described so far also yields feedback gains associated with lateral position and attitude angle of the machine, but these are not shown, since they become redundant when the control is transformed from global to local (see Section 35.7.5). The feedforward gains show convergence to zero as the preview distance increases, which is a long-established feature of preview control [84].

For tight control, the gains are relatively high. In this case, at 10 m/s speed, a view of the road extending only about 17 m forward is all that is necessary. Corresponding necessary preview distances for 40 and 70 m/s are 115 and 256 m, respectively. In terms of time, the necessary preview for full performance increases significantly with speed. The most useful information in the path ahead is about 8 m in front of the motorcycle at 10 m/s, 40 m ahead at 40 m/s, and at 70 m ahead at 70 m/s, around 1 s ahead in time, in each case.

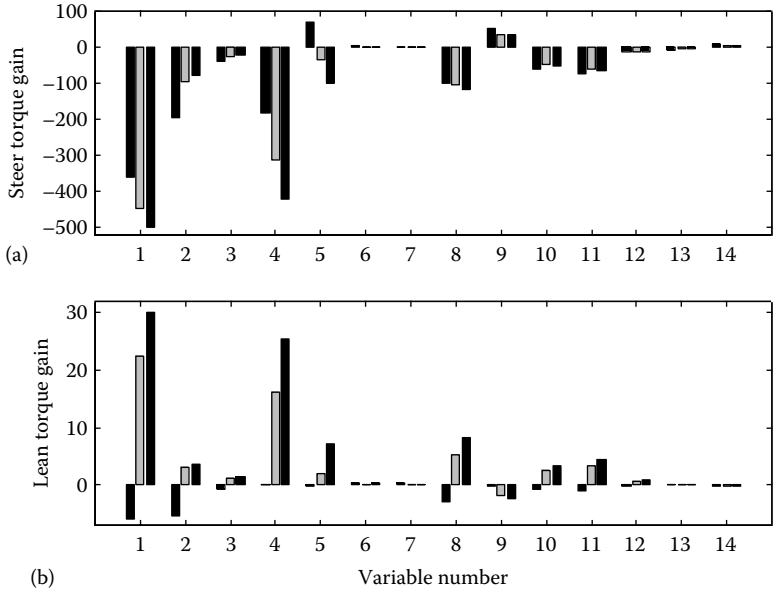


FIGURE 35.62 LQR optimal feedback gains to steering torque (a) and rider body lean torque (b) for $q_1 = 10{,}000$, $r_1 = r_2 = 1$, and speeds of 10, 40, and 70 m/s. In each group of 3 gains, the left-hand one is for 10 m/s, while the right hand one is for 70 m/s. The variables are defined in Table 35.3.

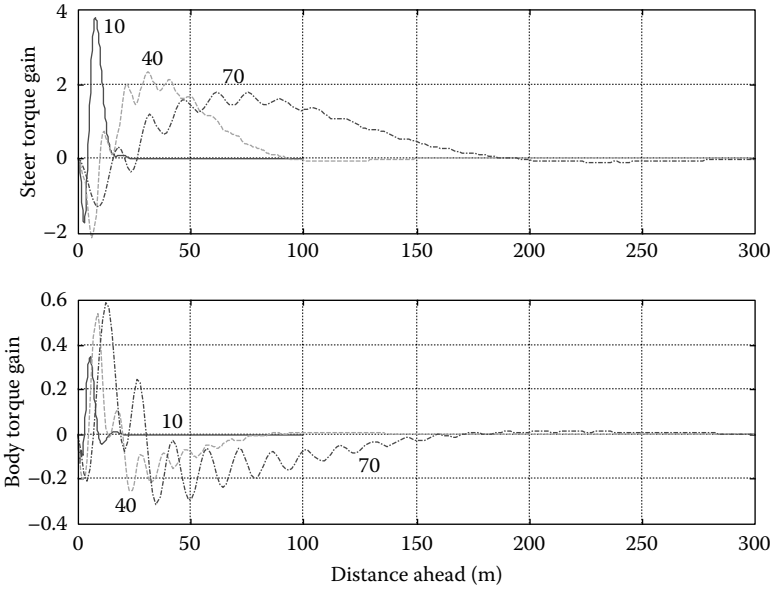


FIGURE 35.63 Preview gain sequences for standard machine system at 10, 40, and 70 m/s in straight running, with $q_1 = 10{,}000$, $r_1 = 1$, $r_2 = 1$.

TABLE 35.3
Variables Associated with LQR Feedback, with Gains Specified in Figures 35.62 and 35.64

Variable #	Variable Description
1	Main frame roll angle
2	Rider hip relative lateral displacement
3	Rider upper body relative roll displacement
4	Frame twist angle
5	Steer angle
6	Rear tire sideslip angle
7	Front tire sideslip angle
8	Main frame reference point lateral velocity
9	Main frame yaw velocity
10	Main frame roll velocity
11	Rider hip lateral velocity relative to the main frame
12	Rider upper body roll velocity relative to the main frame
13	Frame twist velocity
14	Steer velocity

Figures 35.64 and 35.65 show the contrasting results when the control is loose as represented by $q_1 = 100$. The feedback gains are less than half those for the previous tight control, although they follow a similar pattern, with the major gains belonging to roll angle and frame twist angle. The preview gains are reduced very approximately in proportion to the change in the q_1 weighting, and at each speed, the most useful path information is further ahead of the vehicle, at 15, 80, and 170 m for 10, 40, and 70 m/s, respectively, than with the tighter control of Figures 35.62 and 35.63. The inverse dynamics of the motorcycle and rider are apparent in the preview gain sequences, especially for 70 m/s speed. At this high speed, the weave mode of the machine is lightly damped (see Figures 35.18 and 35.19),

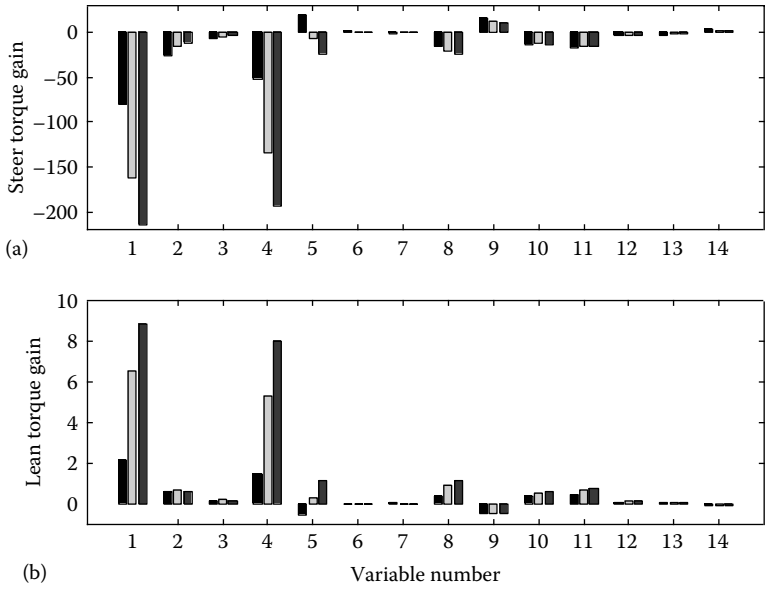


FIGURE 35.64 LQR optimal feedback gains to steering torque (a) and rider body lean torque (b) for q_1 = 100, $r_1 = r_2 = 1$, and speeds of 10, 40, and 70 m/s. In each group of 3 gains, the left-hand one is for 10 m/s, while the right hand one is for 70 m/s. The variables are defined in Table 35.3.

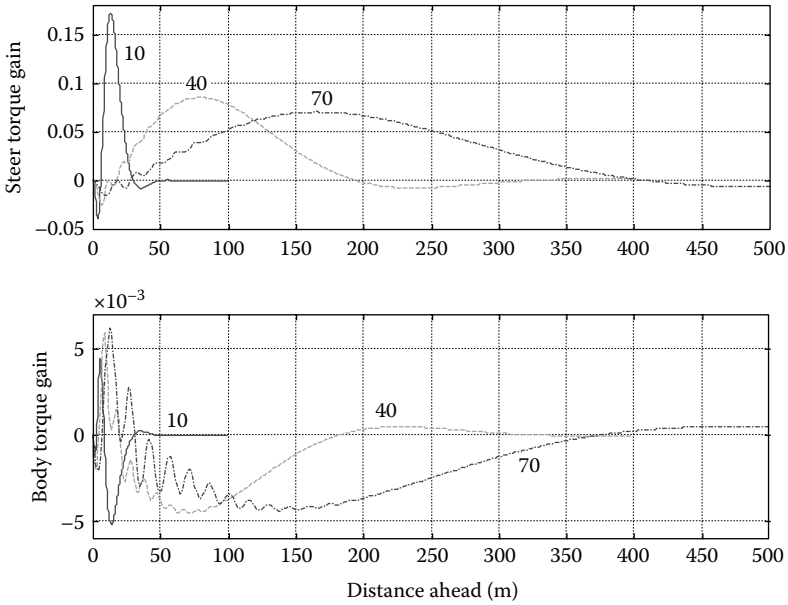


FIGURE 35.65 Preview gain sequences for standard machine system at 10, 40, and 70 m/s in straight running, with q_1 = 100, $r_1 = 1$, $r_2 = 1$.

having a natural frequency of around 4.5 Hz. The oscillations in the preview gain sequence, converted to a time basis, are at this same frequency. This frequency is likely to be well beyond the rider's control bandwidth and it seems unlikely that the mode is usable for control purposes. The main requirement appears to be that the mode is well damped.

With the rewards for additional preview diminishing as the preview increases, it is possible to estimate the preview necessary for more or less full tracking accuracy, as a function of speed,

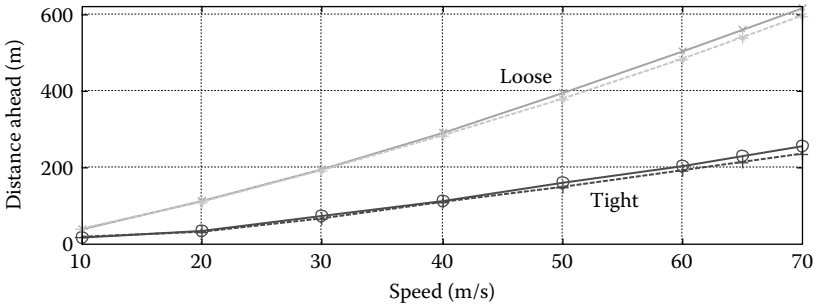


FIGURE 35.66 Preview distances required for "full" performance as functions of speed for standard machine, with tight (q_1 = 10,000) and loose (q_1 = 100) controls and equal weights on the two controls. Each solid line refers to steer torque, while each dashed line refers to body lean torque.

design, and weightings. Indeed, it seems likely that a distinguishing feature of a successful sports motorcycle, designed for responsiveness, is that its need for rider preview of the road is minimized by its design and idea developed for cars in [82]. A simple criterion of the necessary preview is that it is the distance out from the motorcycle at which 98% of the sum of the absolute values of the preview gains is encompassed within that distance. Calculating according to this criterion, for each of the controls, gives the results of Figure 35.66. Substantially, more path preview is necessary for full tracking performance when the control is loose, which result accords with common experience, and that relaxed riding is only possible if one can see the roadway a long way ahead of one's vehicle.

Other preview control law results can be found in [79,80], including some that come from the steering torque as the only control and some that place a lower weighting on the use of rider body torque than on the use of steer torque. The gains relating to steering torque are little influenced by the sort of changes referred to, while, as might be expected, lower weight on body torque leads to corresponding larger gains being found.

35.7.5 Path Tracking Utilizing the Optimal Controls

The simplest approach to path tracking is to use the optimal controls developed in the fixed reference frame as described earlier to close the loop around the motorcycle/road system and to update the state through time by repeatedly multiplying it by the closed-loop state-transition matrix. At each update, a new road sample, from the furthest point along the preview length, enters the problem as input, and the oldest preview data point leaves the problem, since it comes to be aft of the vehicle. This simple approach implies that the restriction to small angles in the linear theory of the control generation carries over into the tracking simulation. Paths that diverge substantially from the starting direction will be followed with imprecision and probably not followed at all if the path direction changes by 90° or more.

A better approach is to transform the problem to a vehicle basis. In such a basis, at the completion of each time step, the reference axes are relocated to align with the motorcycle, and the roadway is described relative to the motorcycle rather than relative to the ground [79–82]. In the repositioning of the reference axes, which are used only to describe the problem and not in any way to govern it, lateral position and attitude angle errors automatically become zero and the previewed road samples are changed from their ground-view to rider-view values. For the transformation to leave the problem unaltered, as it must to make sense physically, those two sets of changes must exactly compensate each other. It is shown in [81] that this is indeed the case. The small-angle theory now requires only that the path direction does not change radically over the preview length. Bearing in mind the relationships between vehicle speed, feasible path curvature, and required preview distance, it can be appreciated that this lesser need is invariably satisfied in practical circumstances.

Each path to be followed is designed to start with a straight section, and the preview states of the combined roadway/motorcycle system are loaded initially with zeros, to correspond with the vehicle having traveled along a straight path, before the simulation commences, for the duration of the preview. The path profile is then "seen" in the distance first, as with a rider with a full view of the road ahead. The simulation proceeds by stepping through time, using the step length 0.02 s employed for the discrete-time optimal control calculations, until a specified number of steps have been completed. At each step, the absolute position and orientation of the motorcycle are updated, and the road data in front of the machine are read in, transformed to the local view of the rider, and used to replace the old values in that part of the system state vector that corresponds to the road profile. Spline interpolation is used to obtain the local lateral road profile at the proper preview points. At the same time, the motorcycle states representing its local lateral position and attitude angle are set to zero. State-feedback and preview gains used are exactly as shown in Figures 35.62 through 35.65.

The maneuvers studied in [80,81] were a lane change and an S-shaped path. Here, we concentrate on a randomly generated path and a sharp 15° direction change maneuver, using 500 preview points in each case. The random path derives from a zero-mean white-noise sequence that is low-pass filtered by a fifth-order Butterworth filter with cutoff frequency 0.25 Hz to make it realistic but interesting [81]. Random path results are shown first, in Figures 35.67 through 35.69 for tight control ($q_1 = 10{,}000$, $r_1 = r_2 = 1$) and 20 m/s and in Figures 35.70 through 35.72 for loose control ($q_1 = 10$, $r_1 = r_2 = 1$) at the same speed. With tight control, the tracking is almost perfect, with the control torques active a little before the roadway starts to curve at about 250 m from the start. Steer and body lean torques are substantially in anti-phase, the latter being much smaller than the former. The maximum roll angle is just over 30° and the rider upper body lean angle is about one-tenth of the roll angle. The maximum steering torque is a comfortable 15 N m. Control torques and responses are a little smaller for loose control and there is some corner cutting in evidence.

Results for the direction change at 20 m/s are shown in Figures 35.73 through 35.75, generated for tight controls, and in Figures 35.76 through 35.78 for loose controls. Precise tracking of the direction change would require high-frequency control inputs. Such a maneuver consequently distinguishes well between tight and loose control. With reduced cost associated with tracking errors,

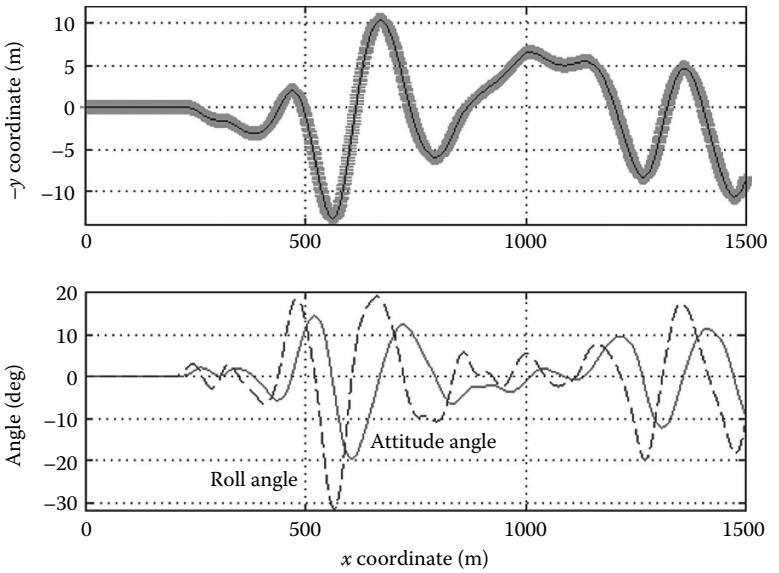


FIGURE 35.67 Random path-tracking results for the standard motorcycle/rider system with tight control ($q_1 = 10{,}000$, $r_1 = r_2 = 1$) at 20 m/s. In the upper part, the roadway is depicted by crosses, while the motorcycle path is shown by the line. The path is defined by sample values at 1 m intervals, obtained by low-pass filtering a zero-mean random sequence with a fifth-order Butterworth filter with cutoff frequency 0.25 Hz.

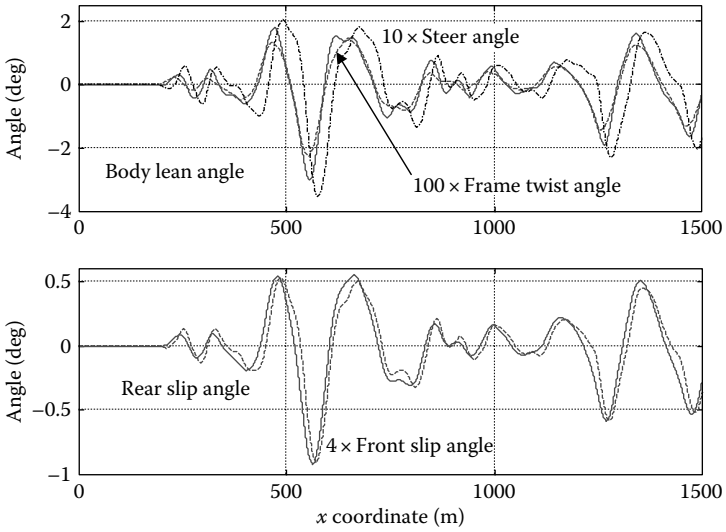


FIGURE 35.68 Random path-tracking results for the standard motorcycle/rider system with tight control at 20 m/s, from the same run as shown in Figure 35.67. The rider's upper body relative roll angle is shown by a solid line, the scaled frame twist angle is dashed, and the scaled steer angle is dash dotted.

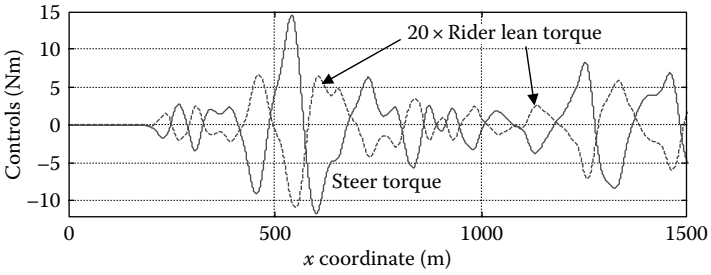


FIGURE 35.69 Control inputs for random path tracking, from the same run as shown in Figure 35.67.

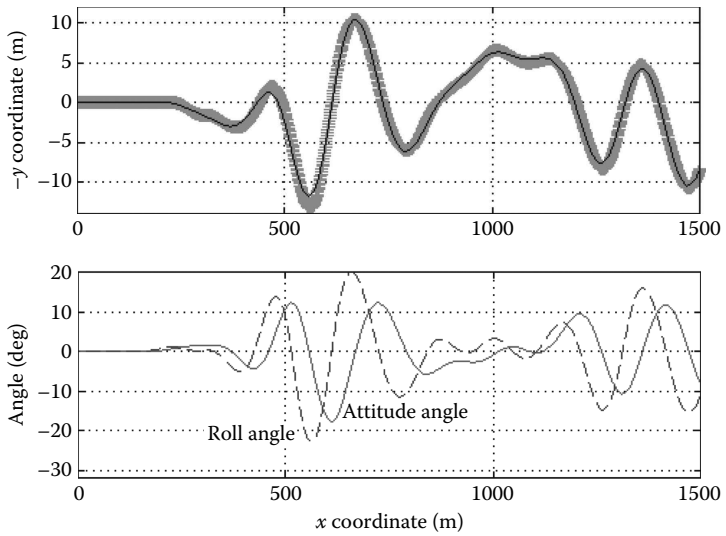


FIGURE 35.70 Tracking the same path as in Figures 35.67 through 35.69 with the standard motorcycle/rider system with loose control ($q_1 = 10$, $r_1 = r_2 = 1$) at 20 m/s. The notation is similar to that of Figure 35.67.

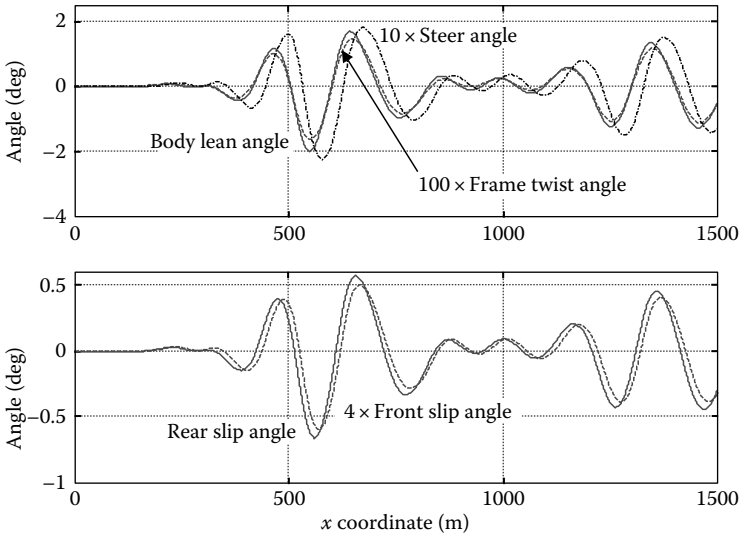


FIGURE 35.71 Tracking the same path as in Figures 35.67 through 35.69 with the standard motorcycle/rider system with loose control ($q_1 = 10$, $r_1 = r_2 = 1$) at 20 m/s. The notation is similar to that of Figure 35.68.

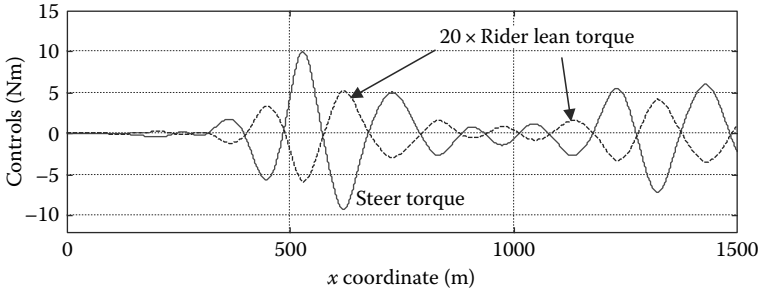


FIGURE 35.72 Control inputs for random path tracking, from the same run as shown in Figure 35.70.

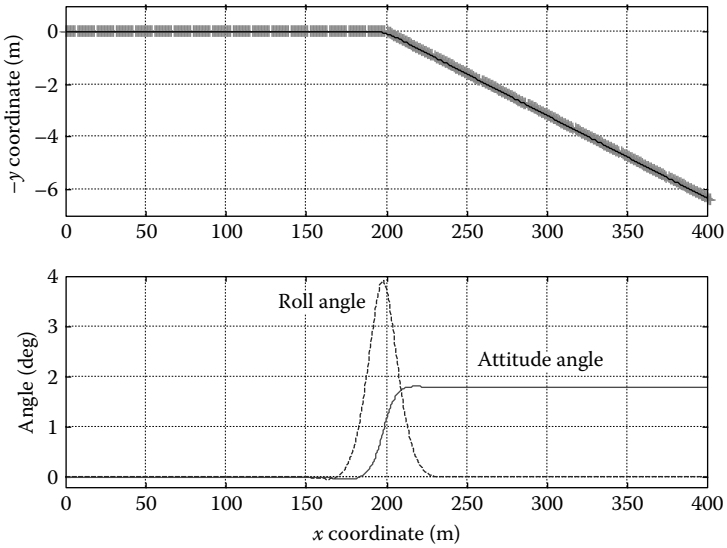


FIGURE 35.73 Path and motion results for 15° direction change maneuver at 20 m/s with tight control ($q_1 = 10{,}000$, $r_1 = r_2 = 1$).

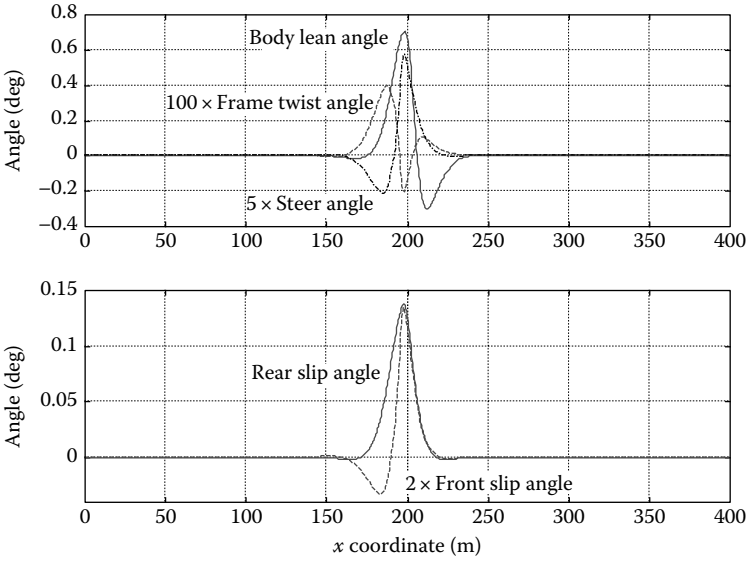


FIGURE 35.74 Motion results and tire sideslip angles for direction change maneuver at 20 m/s with tight control ($q_1 = 10{,}000$, $r_1 = r_2 = 1$).

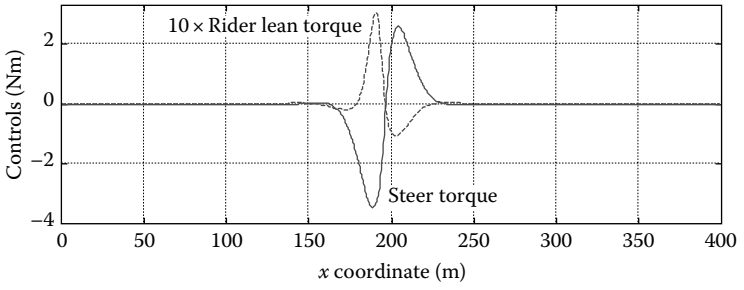


FIGURE 35.75 Control torques for direction change maneuver at 20 m/s with tight control ($q_1 = 10{,}000$, $r_1 = r_2 = 1$).

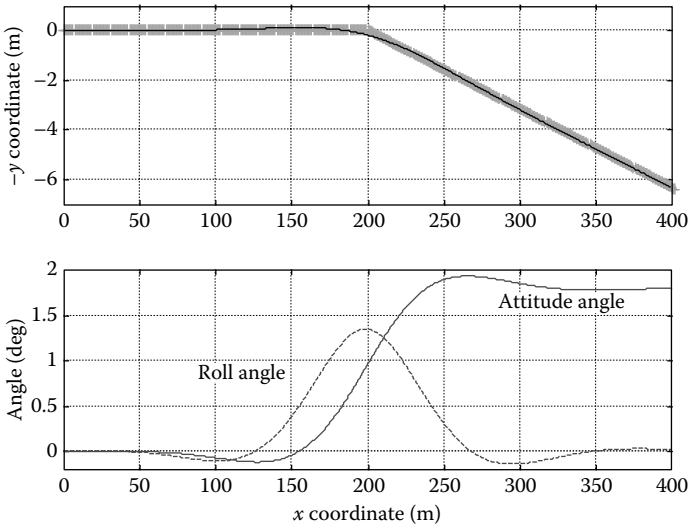


FIGURE 35.76 Path and motion results for 15° direction change maneuver at 20 m/s with loose control ($q_1 = 10$, $r_1 = r_2 = 1$).

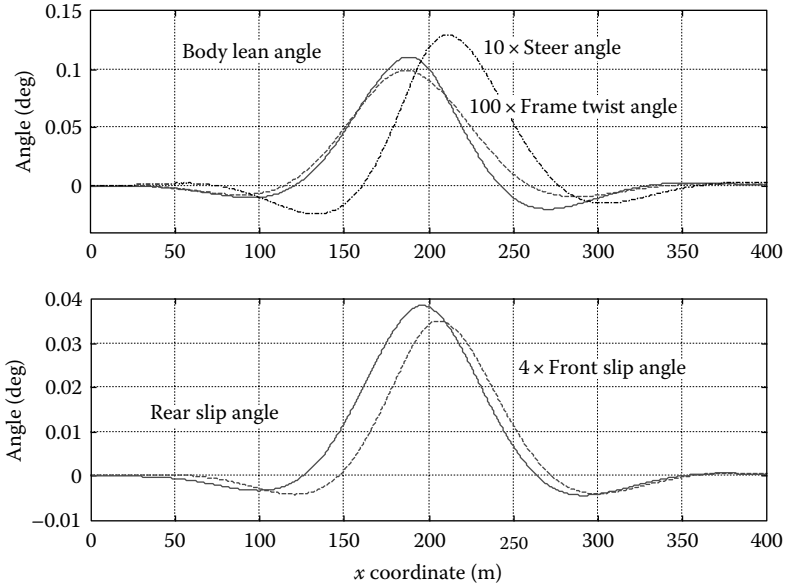


FIGURE 35.77 Control torques for direction change maneuver at 20 m/s with loose control ($q_1 = 10$, $r_1 = r_2 = 1$).

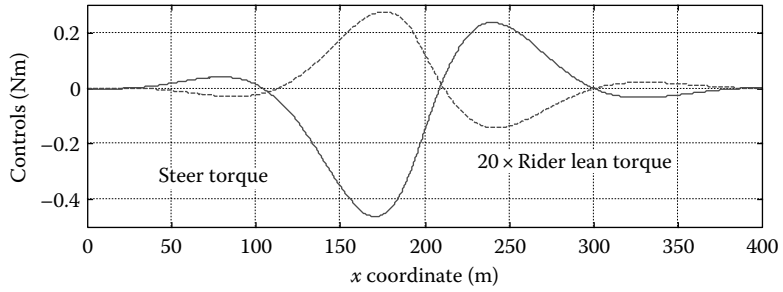


FIGURE 35.78 Control torques for direction change maneuver at 20 m/s with loose control ($q_1 = 10$, $r_1 = r_2 = 1$).

more corner cutting occurs and the control inputs are applied further ahead of the road features. The control inputs themselves are much reduced in magnitude but are spread over a longer time. Peak values of the motion variables are much reduced. For the most accurate tracking, the pre-direction-change countersteering is left late and the front tire is made to develop substantial side force by sideslip. Rider lean control still provides only a small influence on the motions, but this influence is greater than for more relaxed riding. The theory is naturally representing the characteristics of the expert rider and the novice as revealed experimentally by [30]. Also, the theory is illustrating how the rider is able to use his or her knowledge of the motorcycle dynamics to determine an effective preview control plan. As the machine dynamics change due to speed or road condition changes, the rider clearly has a need to alter the plan in sympathy, for the control to remain effective.

35.8 CONCLUSIONS

This contribution has been based on the view that motorcycle handling dynamics is a subset of a man–machine problem, in which the rider is an important element. The motorcycle is a great deal simpler than the rider and we can know about it with much greater certainty. In principle, the motorcycle can be modeled and its behavior can be predicted with some accuracy, provided that the parametric data are sufficient.

The whole topic has been divided into chapters covering rider properties, dynamic theory, modeling needs, behavior, sensitivities of behavioral changes to variations in parameters, and rider control. The behavior considered has included steering-fixed and steering-free operation, steady-turning equilibrium, linear stability and modal properties for straight running and cornering, frequency responses to steer torque, body lean torque and road vertical profile forcing, nonlinear responses to road forcing, and the influences of acceleration and deceleration. Weave and wobble modes are important in the discussion. The value of steering dampers in respect to the wobble and their bad influence on the weave are shown, and a more general steering compensator is mentioned. Force, moment, and power balancing are put forward as valuable processes in the proving of complex multibody models. Quite new work on rider control with full road preview is included. It has been shown how one general and advanced system model can be used as a vehicle for generating a wide range of behaviors.

Research on passive and active steering compensators is ongoing and studies of rider control and handling qualities are far from complete. These appear to be the most interesting and potentially fruitful areas for future investigations. A definitive wisdom on how riders of different stature contribute to the motorcycle structural dynamics has not yet emerged.

ACKNOWLEDGMENTS

Support for the author's work by the Engineering and Physical Sciences Research Council is gratefully acknowledged. Also, the writer has gained much from his discussions on the various topics covered with his colleagues David Limebeer and Simos Evangelou.

REFERENCES

1. Döhring, E., Steering wobble in single-track vehicles, *ATZ*, **58**(10), 1956, 282–286.
2. Eaton, D. J., Man-machine dynamics in the stabilisation of single-track vehicles, Doctoral dissertation, University of Michigan, Ann Arbor, MI, 1973.
3. Watanabe, Y. and Yoshida, K., Motorcycle handling performance for obstacle avoidance, *2nd International Congress on Automotive Safety*, San Francisco, CA, 1973.
4. Jennings, G., *A Study of Motorcycle Suspension Damping Characteristics*, SAE 740628, New York, 1974.
5. Roe, G. E. and Thorpe, T. E., A solution of the low speed wheel flutter instability in motorcycles, *J. Mech. Eng. Sci*., **18**(2), 1976, 57–65.
6. Verma, M. K., Theoretical and experimental investigation of motorcycle dynamics, Doctoral dissertation, University of Michigan, Ann Arbor, MI, 1978.
7. McKibben, J. S., *Motorcycle Dynamics—Fact, Fiction and Folklore*, SAE 780309, Warrendale, PA, 1978.
8. Aoki, A., *Experimental Study on Motorcycle Steering Performance*, SAE 790265, Warrendale, PA, 1979.
9. Weir, D. H. and Zellner, J. W., *Experimental Investigation of the Transient Behavior of Motorcycles*, SAE 790266, Warrendale, PA, 1979.
10. Otto, W. M., Effect of motorcycle accessories on stability, *Proceedings of International Motorcycle Safety Conference*, Motorcycle Safety Foundation, Linthicum, MD, 21090, 1980, pp. 1560–1581.
11. Thomson, B. and Rathgeber, H., Automated systems used for rapid and flexible generation of system models exemplified by a verified passenger car and a motorcycle model, *The Dynamics of Vehicles on Roads and on Railway Tracks* (J. Karl Hedrick ed.), Swets and Zeitkinger, Lisse, the Netherlands, 1984, pp. 645–654.
12. Takahashi, T., Yamada, T., and Nakamura, T., *Experimental and Theoretical Study of the Influence of Tires on Straight Running Motorcycle Weave Response*, SAE 840248, Warrendale, PA, 1984.
13. Nishimi, T., Aoki, A., and Katayama, T., Analysis of straight running stability of motorcycles, *10th International Technical Conference on Experimental Safety Vehicles*, Oxford, U.K., July 1985, 33pp.
14. Hasegawa, A., Analysis of controllability and stability of motorcycles—Analysis of stability at high speed driving, *10th International Technical Conference on Experimental Safety Vehicle*s, Linthicum, MD, July 1985, pp. 480–500.
15. Bayer, B., Flattern und Pendeln bei Krafträdern, *Automobil_Industrie*, **2**, 1988, 193–197.

16. Sugizaki, M. and Hasegawa, A., *Experimental Analysis of Transient Response of Motorcycle Rider Systems*, SAE 881783, Warrendale, PA, 1988.
17. Iffelsberger, L., Application of vehicle dynamics simulation in motorcycle development, *Safety Environment Future* (R. Brenndicke ed.), Forschungshefte Zweiradsicherheit, Vol. 7, 1991, pp. 511–525.
18. Wisselmann, D., Iffelsberger, L,. and Brandlhuber, B., Einsatz eines Fahrdynamik-simulationsmodells in der Motorradentwicklung bei BMW, *ATZ*, **95**(2), 1993, 56–63.
19. Cossalter, V., da Lio, M., Lot, R., and Fabbri, L., Simulation and performance evaluation of race motorcycle dynamics based on parts of real circuit, *Proceedings of International Conference Power Two Wheels*, Pisa, Italy, 1998, SAE 987012.
20. Berritta, R., Biral, F., and Garbin, S., Evaluation of motorcycle handling with multibody modelling and simulation, *Proceedings of 6th International Conference High Tech Engines and Cars*, Modena, Italy, 2000, 11pp.
21. Bortoluzzi, D., Doria, A., Lot, R., and Fabbri, L., Experimental investigation and simulation of motorcycle turning performance, Sicherheit Umwelt Zukunft 111, *Tagungsband der 3 internationalen motorradkonferenz 2000*, Institut für zweiradsicherheit, Essen.
22. Sharp, R. S., Stability, control and steering responses of motorcycles, *Vehicle Syst. Dyn.*, **35**(4–5), 2001, 291–318.
23. Cossalter, V., Da Lio, M., Biral, F., and Fabbri, L., *Evaluation of Motorcycle Maneuverability with the Optimal Maneuver Method*, SAE 983022, 1998. Society of Automotive Engineers, Warrendale, PA.
24. Pick, A. and Cole, D. J., Neuromuscular dynamics and the vehicle steering task, *The Dynamics of Vehicles on Roads and on Tracks* (M. Abe ed.), *Vehicle Syst. Dyn.* (Suppl.) **41**, Taylor & Francis, London, U.K., pp. 182–191.
25. Sharp, R. S. and Limebeer, D. J. N., On steering wobble oscillations of motorcycles, *Proc. IMechE, Part C, J. Mech. Eng. Sci.*, **218**, 2004, 1449–1456.
26. Casanova, D., On minimum time vehicle manoeuvring: The theoretical optimal lap, PhD thesis, Cranfield University, Bedford, U.K., 2000.
27. Gurney, K., *Introduction to Neural Networks*, Routledge, New York, 1997.
28. Sharp, R. S., Some contemporary problems in road vehicle dynamics, *Proc. IMechE, J. Mech. Eng. Sci.*, **214**(C1), 2000, 137–148.
29. Cossalter, V., *Motorcycle Dynamics*, Race Dynamics, Greendale, WI, 2002.
30. Rice, R. S., *Rider Skill Influences on Motorcycle Maneuvering*, SAE 780312, *Motorcycle Dynamics and Rider Control*, SAE SP-428, Warrendale, PA, 1978.
31. Kane, T. R. and Levinson, D. A., *Dynamics: Theory and Applications*, McGraw-Hill, New York, 1985.
32. Leith, D. J. and Leithead, W. E., Survey of gain-scheduling analysis and design, *Int. J. Control*, **73**(11), 2000, 1001–1025.
33. Pacejka, H. B., *Tyre and Vehicle Dynamics*, Butterworth Heinemann, Oxford, U.K., 2002.
34. Koenen, C. and Pacejka, H. B., Influence of frame elasticity, simple rider body dynamics and tyre moments on free vibrations of motorcycles in curves, *The Dynamics of Vehicles on Roads and on Tracks* (A. Wickens ed.), Swets and Zeitlinger, Lisse, the Netherlands, 1982, pp. 53–65.
35. Newland, D. E., *Mechanical Vibration Analysis and Computation*, Longman, Harlow, U.K., 1989.
36. Limebeer, D. J. N., Sharp, R. S., and Evangelou, S., The stability of motorcycles under acceleration and braking, *Proc. IMechE., J. Mech. Eng. Sci.*, **215**(C9), 2001, 1095–1109.
37. Sharp, R. S. and Alstead, C. J. The influence of structural flexibilities on the straight running stability of motorcycles, *Vehicle Syst. Dyn.*, **9**(6), 1980, 327–357.
38. Spierings, P. T. J., The effects of lateral front fork flexibility on the vibrational modes of straight-running single track vehicles, *Vehicle Syst. Dyn.*, **10**(1), 1981, 37–38.
39. Sharp, R. S., The influence of frame flexibility on the lateral stability of motorcycles, *Proc. IMechE, Part C, J. Mech. Eng. Sci.*, **16**(2), 1974, 117–120.
40. Sharp, R. S., The stability and control of motorcycles, *Proc. IMechE, Part C, J. Mech. Eng. Sci.*, **13**(5), 1971, 316–329.
41. Sharp, R. S. and Limebeer, D. J. N., A motorcycle model for stability and control analysis, *Multibody Syst. Dyn.*, **6**(2), 2001, 123–142.
42. Mousseau, C. W., Sayers, M. W., and Fagan, D. J., Symbolic quasi-static and dynamic analyses of complex automobile models, *The Dynamics of Vehicles on Roads and on Railway Tracks* (G. Sauvage ed.), Swets and Zeitkinger, Lisse, the Netherlands, 1992, pp. 446–459.
43. Schiehlen, W. (ed.), *Multibody Systems Handbook*, Springer Verlag, Berlin, Germany, 1990.
44. de Vries, E. J. H. and Pacejka, H. B., Motorcycle tyre measurements and models, *The Dynamics of Vehicles on Roads and on Tracks* (L. Palkovics ed.), *Vehicle Syst. Dyn.* (Suppl.), **29,** 1998, 280–298.

45. de Vries, E. J. H. and Pacejka, H. B., The effect of tyre modeling on the stability analysis of a motorcycle, *Proceedings of Advanced Vehicle Control 1998*, Nagoya, Japan, SAE of Japan, 1998, pp. 355–360.
46. Tezuka, Y., Ishii, H., and Kiyota, S., Application of the magic formula tire model to motorcycle maneuverability analysis, *JSAE Review*, **22**, 2001, 305–310.
47. Sharp, R. S., Evangelou, S., and Limebeer, D. J. N., Advances in the modelling of motorcycle dynamics, *Multibody Syst. Dyn*., **12**(3), 2004, 251–283.
48. Sharp, R. S., Evangelou, S., and Limebeer, D. J. N., Multibody aspects of motorcycle modelling, *Advances in Computational Multibody Systems* (J. Ambrósio ed.), Springer-Verlag, Dordrecht, the Netherlands, 2005, pp. 45–68.
49. http://www.carsim.com
50. Limebeer, D. J. N., Sharp, R. S., and Evangelou, S., Motorcycle steering oscillations due to road profiling, *Proc. Am. Soc. Mech. Eng., J. App. Mech*., **69**(6), 2002, 724–739.
51. Housner, G. W. and Hudson, D. E., *Applied Mechanics—Dynamics*, 2nd edn., Van Nostrand, Princeton, NJ, 1959.
52. Cossalter, V., Doria, A., and Lot, R., Steady turning of two-wheeled vehicles, *Vehicle Syst. Dyn*., **31**(3), 1999, 157–181.
53. Roe, G. E. and Thorpe, T. E., Experimental investigation of the parameters affecting the castor stability of road wheels, *J. Mech. Eng. Sci*., **15**(5), 1973, 365–369.
54. Bandel, P. and Di Bernardo, C., A test for measuring the transient characteristics of tires, *Tire Science and Technology*, Vol. 17, No. 2, 1989, 126–137.
55. Sharp, R. S., Communication to *J. Mech. Eng. Sci*., Proc. I. Mech. E. Part C, **17**(3), 1975, 177.
56. Evangelou, S., Limebeer, D. J. N., Sharp, R. S., and Smith, M. C., Steering compensation for high-performance motorcycles, *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 14–17, 2004, pp. 749–754.
57. Evangelou, S., Limebeer, D. J. N., Sharp, R. S., and Smith, M. C., Mechanical steering compensators for high-performance motorcycles, *ASME J. Appl. Mech*., Vol. 74, 2007, pp. 332–346.
58. Sharp, R. S. and Alstead, C. J., *Frequency Responses of Motorcycles to Steering Torque Inputs and to Front Wheel and Tyre Imperfections*, Road Vehicle Handling, Mechanical Engineering Publication, London, U.K., 1983, pp. 193–200.
59. Bridges, P. and Russell, J. B., The effect of topboxes on motorcycle stability, *Vehicle Syst. Dyn*., **16**(5–6), 1987, 345–354.
60. Nayfeh, A. H. and Mook, D. T., *Nonlinear Oscillations*, Wiley Interscience, New York, 1979.
61. Shaeri, A., Limebeer, D. J. N., and Sharp, R. S., Nonlinear steering oscillations of motorcycles, *Proceedings 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 14–17, 2004, pp. 773–778.
62. Sharp, R. S., Limebeer, D. J. N., and Shaeri, A., Nonlinear vibrations of motorcycles forced by regular road roughness, *Proceedings of the 5th Euromech Nonlinear Dynamics Conference*, Eindhoven, the Netherlands, August 7–12, 2005.
63. Wakabayashi, T. and Sakai, K., Development of electronically controlled hydraulic rotary steering damper for motorcycles, *International Proceedings of the Safety Conference,* IFZ, Munich, Germany, 2004, pp. 1–22.
64. Evangelou, S., Limebeer, D. J. N., Sharp, R. S., and Smith, M. C., An H-inf loop-shaping approach to steering control for high-performance motorcycles, in *Control of Uncertain Systems*, LNCIS 329 (B. A. Francis, M. C. Smith and J. C. Willems eds), Springer-Verlag Berlin Heidelberg, 2006, pp. 257–275.
65. Smith, M. C., Synthesis of mechanical networks: The inerter, *IEEE Trans. Automat. Control*, **47**(10), 2002, 1648–1662.
66. Glover, K. and McFarlane, D. C., Robust stabilization of normalized coprime factor plant descriptions with H-infinity bounded uncertainty, *IEEE Trans. Automat Control*, 3**4**(8), 1989, 821–830.
67. McFarlane, D. C. and Glover, K., Robust controller design using normalized coprime factor plant descriptions, *Lecture Notes in Control and Information Sciences*, **138**, Springer-Verlag, Dordrecht, the Netherlands, 1990.
68. McFarlane, D. C. and Glover, K., A loop-shaping design procedure using H-infinity synthesis, *IEEE Trans. Automat. Control*, **37**(6), 1992, 759–769.
69. Code, K., A twist of the wrist (II)—The basics of high performance motorcycling, Technical Report, California Superbike School, Venice, CA, 1993.
70. Weir, D. H., Motorcycle handling dynamics and rider control and the effect of design configuration on response and performance, Doctoral dissertation, UCLA, Los Angeles, CA, 1972.
71. Weir, D. H. and Zellner, J. W., Lateral-directional motorcycle dynamics and rider control, SAE 780304, in *Motorcycle Dynamics and Rider Control*, SAE SP-428, Warrendale, PA, 1978.

72. Lenkeit, J., *A Servo Rider for the Automatic and Remote Path Control of a Motorcycle*, SAE 950199, New York, 1995.
73. Katayama, T., Aoki, A., and Nishimi, T., Control behaviour of motorcycle riders, *Vehicle Syst. Dyn.*, **17**(3), 1988, 211–229.
74. Katayama, T., Nishimi, T., Okayama, T., and Aoki, A., A simulation model for motorcycle rider's control behaviours, *Proceedings of Small Engine Technology Conference*, Yokohama, Japan, SAE of Japan, 1997, pp. 373–376.
75. Fujii, S., Development of robust rider controlling model for scooter, *Proceedings of Advanced Vehicle Control '04*, Appeldoorn, Arnhem, the Netherlands, 6pp.
76. Frezza, R., Beghi, A., and Saccon, A., Model predictive and hierarchical control for path following with motorcycles: Application to the development of the pilot model for virtual prototyping, *43rd IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, December 2004, pp. 767–772.
77. MacAdam, C. C., Understanding and modelling the human driver, *Vehicle Syst. Dyn.*, **40**(1–3), 2003, 101–134.
78. Watanabe, Y., Private communication, December 2005.
79. R. S. Sharp, Optimal linear time-invariant preview steering control for motorcycles, *The Dynamics of Vehicles on Roads and on Tracks* (S. Bruni and G. Mastinu eds), supplement to VSD 44, Taylor and Francis (London), 2006, 329–340.
80. Sharp, R. S., Motorcycle steering control by road preview, *Trans. ASME, J. Dyn. Sys., Meas. Control*, Vol. 129, 2007, pp. 373–381.
81. Sharp, R. S. and Valtetsiotis, V., Optimal preview car steering control, *ICTAM Selected Papers from 20th International Congress* (P. Lugner and K. Hedrick eds), supplement to VSD 35, 2001, pp. 101–117.
82. Sharp, R. S., Driver steering control and a new perspective on car handling qualities, *Proc. IMechE, Part C, J. Mech. Eng. Sci.*, **219**, 2005, 1041–1051.
83. Anderson, B. D. O. and Moore, J. B., *Linear Optimal Control*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1971.
84. Tomizuka, M., Optimal linear preview control with application to vehicle suspension – revisited, *Trans. ASME, Journal of Dynamic Systems, Measurement and Control*, **98**, 1976, 309–315.

Part VI

Man–Vehicle Interaction

36 Vehicle Comfort

Karl Siebertz

CONTENTS

36.1 Comfort and Discomfort 1379
36.1.1 Definition and Theories 1380
36.1.2 Aspects of Comfort 1380
36.1.3 Ergonomic Rules and Laws 1381
List of Symbols 1382
36.2 Human Body 1382
36.2.1 Body Size 1382
36.2.2 Resonance Frequencies 1383
36.2.3 Human Locomotor System 1384
36.2.4 Heat Balance 1385
36.3 Comfort Evaluation 1386
36.3.1 Subjective Rating 1386
36.3.2 Test Methods 1386
36.3.2.1 Human Motion 1386
36.3.2.2 Seat Vibrations 1387
36.3.2.3 Seat Pressure Distribution 1388
36.3.2.4 Electromyography 1388
36.3.2.5 Climate 1389
36.3.2.6 Others 1390
36.3.3 CAE Models 1390
36.3.3.1 Kinematic Models 1390
36.3.3.2 Inverse Dynamic Models without Muscles 1390
36.3.3.3 Inverse Dynamics Models Including Muscles 1390
36.3.3.4 Forward Dynamic Models to Simulate the Human Body Impedance 1390
36.3.3.5 Forward Dynamic Models to Simulate the Vibration of the Human Body 1391
36.3.3.6 FEM Prediction of the Seat Pressure Distribution 1391
36.4 Conclusion 1391
Acknowledgments 1391
References 1392

36.1 COMFORT AND DISCOMFORT

Vehicle comfort is a known customer want and receives increasing attention as a strong sales argument. Over the past decades, several factors have changed: The basic function of cars is no longer a discriminator between the brands. Customers expect a higher level of driving quality even for smaller cars, sports, or off-road vehicles. The average customer spends more time in the car. Higher traffic density and speed increase the driver's workload. The average age of the driving population is increasing, at least in Europe [28] and the United States. Additionally,

many vehicles are used as working place. All of the aforementioned facts are good arguments for cars with increased level of comfort.

36.1.1 Definition and Theories

What is comfort? Here are a few definitions found in lexica and the literature:

Merriam–Webster [17]: "Pronunciation: \'kəm(p)-fərt\ Function: transitive verb Etymology: Middle English, from Old French conforter, from Late Latin confortare to strengthen greatly, from Latin com- + fortis strong"

Wikipedia [37] (translated): "Comfort can be described as the absence of discomfort. Discomfort means noticeable unpleasant sensations. Basic assumption is an ongoing comparison between actual sensation and previous experiences. The current situation will not be considered as unpleasant as long as there is no discrepancy between sensation and expectation for the specific context. Deviations will be noticed, which means that comfort depends on the expectations of the observer. …" Zhang et al. [40] proposed a 2D scale. Positive aspects of "liking" are associated with comfort. Negative aspects of "suffer" are associated with discomfort. Comfort and discomfort can be present at the same time. Bubb [4] also promotes the 2D scale and in addition suggests a general sequence for the discomfort aspects.

"Absence of discomfort" does not really provide a guideline to create a comfortable vehicle design. The 2D scale can explain why customers are willing to accept a certain level of discomfort and still like a vehicle, but it is very important to separate comfort from other positive attributes such as quality impressions [11] and performance feel. One attempt using a positive definition could be: "Comfort is the result of correct ergonomic design and additional convenience features." The correct ergonomic design will avoid discomfort, and the convenience features are able to create positive emotions that can be associated with comfort. Even with this short definition, comfort will remain multidimensional and very subjective.

The definition in Wikipedia includes a threshold theory that has been widely observed. If the unpleasant sensation exceeds a certain limit, it will become a discomfort. Different aspects of comfort are present at the same time, and the occupants of the vehicle will combine their individual sensations to a single score that is mainly determined by the worst comfort discipline. This is the "bad tooth theorem." If one tooth hurts, the typical patient will not be in the mood to consider 27 good teeth and come up with an average pain score. Recent health surveys in Europe indicate an increasing prevalence of back pain [3]. It is very likely that this kind of health problems will shift the priorities of the consumer toward comfort.

Coming back to the vehicle, this results in several consequences:

1. All comfort aspects must be considered simultaneously.
2. A comfortable vehicle has a good performance in every comfort discipline.
3. Designing for comfort requires deep knowledge of the human physiology.

36.1.2 Aspects of Comfort

The comfort aspects work toward three main functions: support, protect, and delight.

Based on the classical definition of ergonomics, vehicle design should support the user and make the driving as easy as possible. This includes a proper seating position with postural support under all conditions, a sufficient range of motion, a well-balanced pressure distribution, logically clustered and readable instruments, easy to reach steering and pedals, operational forces in the right range, and large enough door openings. However, the expectation of the passenger can be significantly different from the driver. While the driver needs to be "connected" to the vehicle in order to fulfill the driving task, most of the passengers tend to put more emphasis on the smoothness of the ride. Yet in both cases, sitting in the vehicle is just one of the possible situations. A comfortable vehicle supports all operations within the normal range of usage, for instance, loading/unloading, drinking coffee, or reading a map.

Driving a vehicle implies all kinds of exposures. Noise is generated in the engine compartment, through the tires, the wind, and from other traffic. Vibrations are present over a wide range of frequencies. Sudden impacts generate inertial forces in the body. The occupant needs to be protected from the weather, dirt, and pollution. A comfortable vehicle is essentially a little cozy survival space, a home on wheels.

Like in any other home, a nice atmosphere and entertainment would not hurt. This is absolutely consistent with the Latin roots of the word "comfort" because these are measures to "charge the human battery." From the physiological point of view, "charge" might even be a better word than "delight." Delight features can be used to strengthen a brand image. However, a bunch of features without proper ergonomics does not deliver a comfortable car. The absence of delight features does not create discomfort but might bring it earlier to the attention of the driver.

36.1.3 Ergonomic Rules and Laws

The driver has to watch the traffic, find his or her way, follow the street, control the speed, switch gears, use the brakes, operate controls, and sometimes even perform additional tasks in parallel to remain part of our information society. This is real work. Looking at the vehicle as a workplace, several rules and evaluation criteria can be used in the traditional sense of ergonomics [24]. The human body has not originally been designed to drive a car. In early days, many technical restrictions forced the driver to adapt to unnatural conditions behind the steering wheel. Nowadays, the car can be designed to accommodate the human. A good workplace design always takes into account the balance between capacity and demand [20]. Capacity in this context is the human ability to sustain a certain amount of physical and mental workload. Demand is the required workload to fulfill the specific task. Considering the vehicle as a workplace can explain why subjective comfort perception of the same vehicle can be so different between users.

Comfort requires a positive and stable physical and mental state, which can only be obtained if the demand is lower than the capacity. In a driving situation, the demand can vary significantly, depending on the vehicle design, weather conditions, traffic, speed, road conditions, or additional tasks. The interindividual capacity varies significantly with age, strength, size, weight, intelligence, and education to name a few. Even the intraindividual capacity is subject to a remarkable variation and depends on factors such as health state, sleep, mood, stress, time of the day, or taste of the coffee in the morning. Sometimes an uncomfortable vehicle is perceived as comfortable because it has only been evaluated in situations where the capacity was high and the demand low. Just imagine you own a nice old sports car that you only drive on Sundays when the weather is perfect, the roads are empty, and you have had plenty of sleep and a good breakfast and pick a scenic route. The same car might feel terrible on a dark and foggy Monday morning in heavy traffic under time pressure after a short night.

Many design rules for workplaces can be applied directly to vehicle design. Some ergonomic books are listed in the references [13,22,24]. The biggest difference between vehicles and standard workplaces is the unavoidable vibration of the body. Therefore, vibration exposure in particular should be considered during the vehicle development.

The EU directive 2002/44/EC [8] defines minimum safety and health requirements for workplaces. If a vehicle is used as a workplace, this directive applies for vehicles as well. The vibrations of the body as a whole are measured in accordance with ISO standard 2631 (1997). The hand-arm vibrations are measured according to ISO 5349 (2001). While the hand-arm vibration exposure should not be an issue in vehicles, whole-body vibrations, however, might become relevant under certain circumstances. Unfortunately, the vibration exposure depends on many factors and not all of them belong to the vehicle design. Road condition, driving behavior, human body mass, individual seating position, or muscle tension cannot be prescribed by the development engineer. Looking at the vehicle, vibration damping cannot be assigned to a single component or subsystem. It is an overall vehicle performance requirement, involving tires, wheel size, suspension, vehicle mass, body stiffness, carpet, the seat, and to some extent the steering column. The main drivers are, however, the suspension and seat.

ISO 2631 metrics are root mean square (RMS) or vibration dose value (VDV) of the weighted acceleration during a given time span. Other metrics are reported in Section 31.2.1. Comfort of off-road vehicles is dealt with in Chapters 12 and 34:

$$\mathrm{VDV} = \left[\int_{t=0}^{T} a_w^4(t)dt \right]^{0.25} \quad \text{with } a_w(t) = c(f) \cdot a(t)$$

and

$$c(f) = \begin{matrix} 0.4 \\ f/5.0 \\ 1 \\ 16/f \end{matrix} \quad \text{for} \quad \begin{matrix} 0.5 < f < 2 \text{ Hz} \\ 2 < f < 5 \text{ Hz} \\ 5 < f < 16 \text{ Hz} \\ 16 < f < 80 \text{ Hz} \end{matrix}$$

Further exposure must be prohibited if RMS or VDV exceeds their exposure limit value (ELV) for the daily exposure: $\mathrm{RMS_{ELV}} = 1.15 \frac{\mathrm{m}}{\mathrm{s}^2}$, $\mathrm{VDV_{ELV}} = 21 \frac{\mathrm{m}}{\mathrm{s}^{1.75}}$.

LIST OF SYMBOLS

a	Acceleration
a_w	Weighted acceleration
$c(f)$	Frequency-dependent weighting factor
f	Frequency

36.2 HUMAN BODY

Obviously, it would be far beyond the scope of this chapter to explain, in detail, the human anatomy and physiology. However, vehicle comfort involves the vehicle and the human. Engineers and designers should, therefore, know some basic facts about the human body (see [20,23,25,32]).

36.2.1 Body Size

The first comfort-related development task is the dimensional layout of the vehicle interior, the so-called people package. It can be very difficult to use a car that does not accommodate the user's body size. Potential issues can be visual obstructions, problems to reach pedals or controls, uncomfortable joint angles, and collisions with parts of the interior, such as steering column shroud or headliner. In particular, the seat design is very sensitive to the body dimensions.

In ergonomics, it is common practice to work with percentiles, for instance, the 5th, 50th, and 95th percentiles. The 5th percentile is small; only 5% of the population is smaller. The 50th percentile represents "Joe average," whereas 95% of the population is smaller than the 95th percentile. A design that works well for the entire range between the 5th percentile female and 95th percentile male will probably not create issues in real-world usage. Unfortunately, there is a significant geographical variation of body sizes (Figure 36.1).

Additionally, there is a drift toward larger body sizes of about 1 cm per decade [9]. The reference sizes of the percentiles need to be updated and projected to the planned production period of the vehicle.

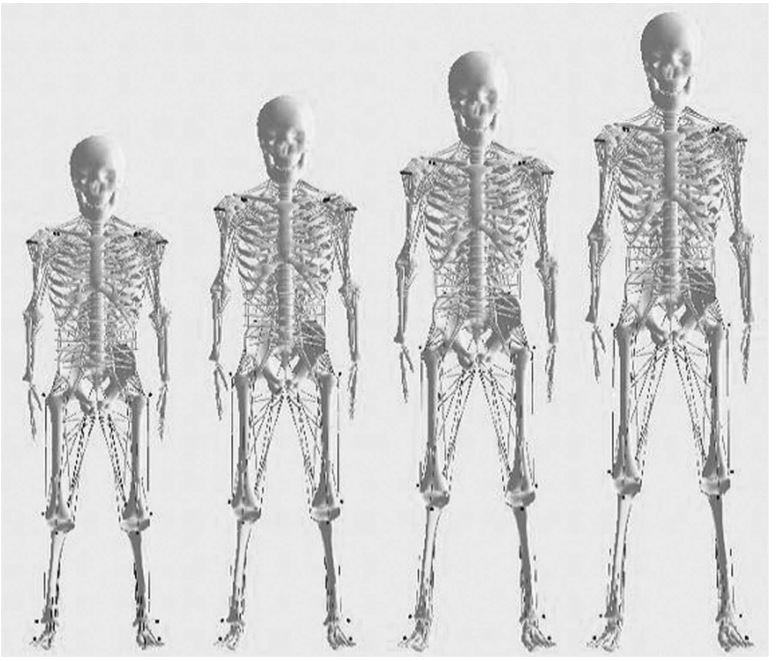


FIGURE 36.1 Scaled model of the human body, 5% female, 50% female, 50% male, and 95% male. (Plot generated by Ford Research and Advanced Engineering, Cologne, Germany.)

36.2.2 Resonance Frequencies

Why are resonance frequencies so important? Resonances of the human body can cause various kinds of discomfort [10] (Table 36.1). In many cases, the customers are not irritated by nor do they even feel the vibration but suffer from the resonance of internal organs. A very critical range lies between 3 and 6 Hz.

There is normally a resonance frequency range for different body parts due to people having different proportions and weight. Obviously, not all people with the same size automatically

TABLE 36.1
Human Body Resonances

Excitation Frequency (Hz)	Human Body Response
3–4	Strong resonance in the cervical vertebrae
3–6	Resonance of the stomach
4	Peak resonance of the lumbar vertebrae
4–5	Resonance in the hands (difficult to make aiming motions)
4–6	Resonance in the heart
5	Very strong resonance in the shoulder girdle
5–20	Resonance in the larynx (voice changes)
5–30	Resonance in the head
10–18	Resonance in the bladder (urge to urinate)
20–70	Resonance in the eyeballs (difficult to see)
100–200	Resonance in the lower jaw

Source: Kroemer, K. and Grandjean, E., *Fitting the Task to the Human*, Taylor & Francis, Philadelphia, PA, 1997.

have the same weight. Similarly, the damping behavior of the vehicle seat depends on weight and proportions of the occupant as well.

36.2.3 Human Locomotor System

Compared to the past, today's usage of the human locomotor system has changed dramatically, too fast for the evolution to follow up with an optimized version of the "Homo sedens" (sitting human). As a result, using our body that has been optimized for the traditional life, dominated by physical activity, we experience discomfort in static seating positions with poorly distributed muscle tension. In many cases, this even leads to significant pain or reduced muscular performance [3]. Sport teachers and gym managers have already noticed that the current lifestyle of younger people has led to reduced average maximum muscle forces.

Therefore, it is important to analyze all driving tasks with respect to the muscular activity. This is a real challenge because the human locomotor system consists of more than 200 bones, connected with various kinds of joints and more than 600 muscles that can be considered as independent actuators [25]. Adding all degrees of freedom in the human skeleton, the total number will be significantly smaller than the number of skeletal muscles. In mathematical terms, this means that the equation system is indeterminate. In physiological terms, this means that the human body has several options to recruit the muscles and still fulfill the same movement task [20]. Many muscles cross two joints, others only one joint. Tendons are in place to optimize the effective lever arms and introduce wrapping points in the line of action, if a simple straight line would not be ideal for the human body. A complex system of ligaments is needed to utilize different bones and allow for the right amount of movement. For instance, the knee joint would almost fall apart without ligaments since the contacting bone surfaces of the femur (long bone of the upper leg) and tibia (long bone of the lower leg) are not congruent.

Typical driving comfort issues related to the locomotor system are:

1. *A lack of postural support in the seat*. The muscles become overloaded after a certain time and cannot stabilize the body any more. This leads to fatigue and higher loads on the passive locomotor system. Problems with cartilage (e.g., vertebral disks) and ligaments are resulting long-term consequences.
2. *Muscular effort is too high to perform the task*. This could, for instance, apply to liftgates or mechanical handbrakes. It is more of an issue for elderly people as they often, simultaneously, have limitations of the joint movement range.
3. *Muscular efforts are too high to fulfill a task over the required time*. This applies for all control tasks such as steering and pedal operation. After just one hour, the maximum muscle force drops down to less than 10% of the short-term maximum force. Therefore, even low activation levels can become critical.
4. *Activation levels are too low or have the wrong characteristics in order to fulfill the control task*. These issues are a little more complicated because the goal is not simply to reduce the effort. The driver is part of the man–machine control system, at least until autopilots become standard in our vehicles. Using the current vehicle layout, the entire control communication (steering, shifting, braking, accelerating) relies on the human locomotor system's input. The control communication is bidirectional: The driver submits commands through mechanical forces or torques, and the vehicle gives feedback through the same control elements (a mathematical formulation can be found in Mitschke [18]). Unfortunately, the driver moves during driving (probably the original reason to use the car) resulting in dynamic loads, on the same extremities that are used to operate the controls.

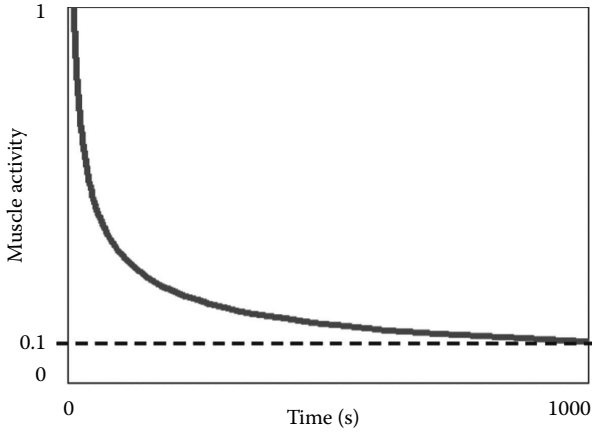


FIGURE 36.2 Muscle fatigue (50% slow-twitch fibers). The maximum force can only be developed for a short time due to metabolic constraints. Slow-twitch fibers are optimized for endurance; fast-twitch fibers are optimized for peak force [15]. (Ford plot using data from Niemi, J. et al., *J. Biomech.* 29, 451, 1996.)

The human locomotor system is fully equipped with biological strain gauges in the muscles, joint position sensors, accelerometers, force sensors in the skin, and a clever kind of "gyro" to trace the head rotation [23]. Without such mechanisms, it would be completely impossible to walk or perform other movement tasks. Recent developments of robots have exposed the difficulty of walking. It took decades until machines could balance themselves on two legs and move forward in a controlled manner. Driving a car takes advantage of the sophisticated control features that our body offers. In a sense, we become a part of a symbiotic man–machine system (Figure 36.2).

36.2.4 Heat Balance

The human body needs to keep the core temperature in a very tight tolerance band. A sophisticated temperature regulation performs this task automatically, using either increased metabolism for heating or sweating/respiration for cooling or blood vessel diameter changes in combination with heart rate changes to concentrate or spread the blood flow within the body. The optimum temperature range is affected by the humidity, differs between individuals, and depends on the clothing worn. For dry air, the temperature range is larger compared to wet air, and women typically like higher temperatures than men. The regulation process is not noticeable until it exceeds a certain level and then it not only becomes noticeable, but it may eventually cause fatigue due to the massive physical effort required. In addition to discomfort, a poor temperature stabilization in the vehicle can even cause sickness or increase the reaction time of the driver.

The heat balance in a vehicle is influenced by many external factors such as ambient temperature, sun radiation, moisture, and wind speed. Compared to regulating the temperature in a house, the task is a little more difficult in the vehicle because the driver produces more heat than the passengers, the cabin air volume is small, the heat capacity of the car interior is low, the heat production of the car itself varies, and the individual body regions are exposed to a different heat transfer situations. Unfortunately, it is difficult for the human body to react locally to different temperatures. For instance, sweat production cannot be controlled individually for the front and rear parts of the body's trunk. Two independent strategies are available for the car designer: (1) Maintain the perfect climate in the car. (2) Support the natural thermoregulation of the human body, for example, the seat covers a significant heat exchange area of the human body and should have a sufficient water vapor permeability.

36.3 COMFORT EVALUATION

36.3.1 Subjective Rating

Subjective rating is the oldest method to evaluate comfort and will probably continue to survive for some time. As mentioned earlier, the human body is fully equipped with hundreds of biological sensors. Experienced evaluators essentially use these sensors in order to detect small differences in design, which are not yet measurable.

Many different evaluation schemes have been reported in the literature, and it is beyond the scope of this chapter to come up with a complete review or guideline. However, some basic rules should be mentioned:

1. *Clear definition of the evaluation criteria and precise description of the evaluation conditions.* Comfort is multidimensional. This means that several independent criteria need to be defined and separated from each other. For each criterion, there should be one critical task or boundary condition where the differences between good and bad design are evident. Many criteria require real driving exercises on the test track. Pure showroom evaluation is not appropriate.
2. *Reference vehicle.* It is much easier to compare two vehicles than rate a single vehicle without any reference. The reference vehicle needs to be in a controlled condition over the time of usage.
3. *Descriptive rating scale.* The rating scale should be clearly defined and not too fine. It is important that all evaluators have the same understanding of a particular rating level.
4. *Customer-oriented weighting.* The multidimensional comfort rating needs to be combined to one total score for each vehicle. A representative picture can only be obtained if all important comfort aspects are captured and weighted according to the customer preference for the particular vehicle type.
5. *Trained evaluators.* Subjective vehicle evaluation is like wine tasting. The evaluator needs much experience, the ability to focus on specific aspects, and the ability to separate the differences caused by boundary conditions from the difference in performance.
6. *Sufficient evaluations.* Even trained evaluators do not always come to the same conclusion. However, a smaller number of observations are needed, compared to regular customers. For a proper comfort evaluation, it is necessary to cover a certain range of the population with a number of selected evaluators because in some comfort aspects like vibration damping, postural, or lateral support, the comfort performance depends on the occupant size.

For design optimization, it can be helpful to use bimodal scales instead of 1D scales. For instance, a seat can be too wide or too narrow. Just rating the seat width on a scale from 1 to 10 would not give guidance on how to improve the design.

36.3.2 Test Methods

Test methods typically focus on one specific comfort aspect, for instance, vibration damping or seat pressure distribution. A complete comfort assessment would require a massive test program. The advantage of tests over a subjective rating is a well-defined output that in many cases can directly be used to detect design weaknesses. However, most of the comfort-related tests involve human test subjects, which is a real drawback in terms of repeatability and reproducibility.

36.3.2.1 Human Motion

Motion analysis has a long history in biomechanics. Originally used in clinical applications such as gait analysis, the same methods have now started to migrate into the vehicle development. Motion analysis is helpful to investigate how people perform operation tasks such as ingress/egress [12]. Depending on vehicle design parameters such as door opening height or step over height, many



FIGURE 36.3 Strain gauge goniometer to measure joint angles. (From Siebertz, K. et al., The strain gauge goniometer, a new sensor to measure dummy joint angles under crash conditions, *SAE World Congress 2000*, in: *Safety Test Methodologies*, SAE 2000, Special Publications, Warrendale, PA, 2000; Photo from Ford Research and Advanced Engineering, Cologne, Germany.)

people change their movement pattern completely. Motion analysis also builds the backbone of further biomechanical analysis with computer-aided engineering (CAE) models. Motion analysis typically works with a simplified rigid body model and tries to replicate the kinematics in terms of joint angles or joint center movement.

The simplest way to record movement is videotaping or a series of digital photos. Markers should be used to identify prominent points, if possible joint locations or bony landmarks [38]. Skin movement and loose clothing are critical. A 3D reconstruction requires at least two independent unobstructed views, which can be difficult to obtain inside the vehicle. A full-blown motion tracking works with a set of four or more synchronized cameras to make sure that at least two views are available at any time. The redundancy also reduces errors due to optical distortion. In the past, many competing systems were on the market, mostly optical, some based on ultrasound, working with active or passive markers. Video-based systems and infrared-based systems with passive markers seem to be preferred.

For single joints, it is always a good option to use goniometers that are taped on the body. They deliver joint angles directly and work rather reliably. Using this technique for the whole body requires a special suit (in all body sizes that are to be investigated) (Figure 36.3).

Small movements of the human spine can be recorded noninvasive with a posture monitor, using ultrasonic sound waves to detect small distance changes between eight sensors that are taped on the skin.

36.3.2.2 Seat Vibrations

Vehicle seats can dampen the car body vibrations quite effectively. It is easy to measure seat track vibrations with regular accelerometers [1,10]. However, seat cushion vibrations depend not only on the seat but also on the occupant. A simple rubber pad with integrated accelerometer can be placed on top of the seat cushion to record the seat cushion accelerations. The rubber pad increases the contact area and is flexible enough to follow the seat shape. Hydro-pulser tests can also be very helpful for seat development. In this case, the seat is mounted on a rig and accelerated according to the measured accelerations from the seat track on the road test. In both cases, the human test subject is used more or less as passive dummy, just to load the seat. Several vibration dummies have been proposed so far to overcome this shortage, but there is no common automotive standard yet.

Most of the studies reported in the literature focus on vertical vibration of the seat pad, and it is common sense that these correlate to a subjective vibration comfort perception. However, it is debated to which extent the seat back and the other directions of the vibration contribute to the total picture. Beyond 30 Hz, the seat does not vibrate as a total system, but there are noticeable local effects, caused by resonances or modal vibrations.

Such comfort evaluations are mainly concerned with the following numerical results: (1) absolute vibration in terms of RMS or VDV and (2) ratio between seat cushion acceleration and seat track acceleration. Obviously, the amount of vibration depends on the road conditions. It is, therefore, important to use well-known roads, for example, on a test track.

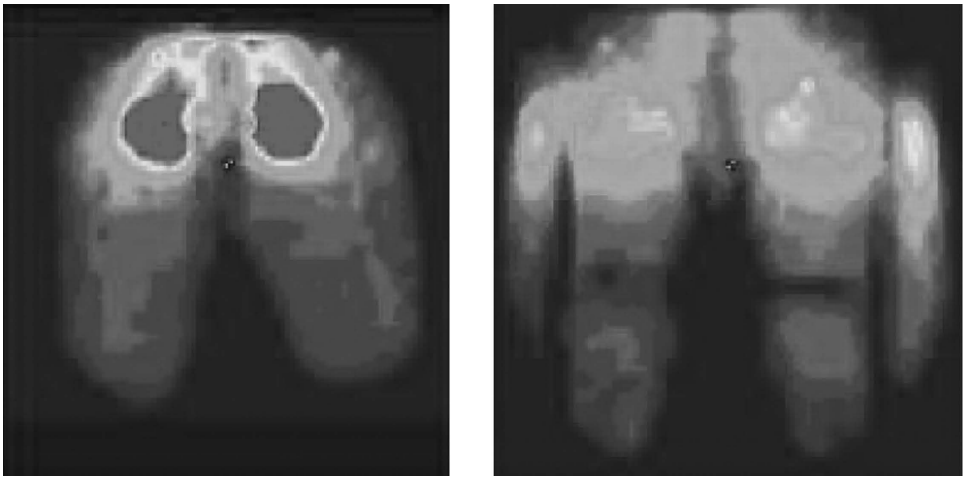


FIGURE 36.4 Seat pressure distribution of a canteen chair (left) vs. a vehicle seat using the same test subject. The canteen chair has been designed (on purpose?) for short-term usage. (Plot generated by Ford Research and Advanced Engineering, Cologne, Germany.)

36.3.2.3 Seat Pressure Distribution

Since the human body is equipped with a large number of pressure sensors, the occupants are very sensitive to the contact pressure. For example, pressure hot spots, high-pressure gradients, or small contact areas cause discomfort [16,33]. There is always a trade-off between body support and uniform pressure distribution. Without a certain pressure, there is no way to transmit the required support forces to the body, especially under dynamic conditions (Figure 36.4).

Many design weaknesses of seats, therefore, can be detected with the measurement of the seat pressure distribution. Unfortunately, the pressure distribution for the same seat can look totally different, depending on the occupant. A seat evaluation, therefore, requires measurements with a variety of people. Seat indenter with special "natural-shaped" padding is under development, but there is no standard yet.

Pressure distribution measurement has been around in biomechanics for decades [30]. An early application was the treatment of diabetic feet with custom-made shoes to optimize the pressure distribution [2]. Modern pressure mapping systems use several hundred pressure sensors per mat, can work with two mats at the same time, and record more than 10 frames per second. A high sampling frequency is helpful for dynamic measurements on test tracks, for example, to check the lateral support or bottoming out on bumpy roads.

36.3.2.4 Electromyography

The muscular force cannot be measured directly, but for many muscles, it is possible to record the electrical activity. Electrodes placed on the skin above the muscle can detect small changes of the electric potential compared to areas further away to the muscles [6]. This is called surface electromyography (EMG). A more precise measurement can be obtained using needle electrodes that are placed directly in the muscle. However, this causes discomfort in itself and, therefore, does not help to evaluate comfort. Surface EMG has a long history in clinical and sports applications because it is the only noninvasive way to obtain test data related to muscular activity. EMG is the indispensable tool to generate the validation data for further CAE studies with biomechanical models (Figure 36.5).

The signal post-processing depends on the goal of the study. For single movements such as pulling a handbrake, the signal would be rectified and filtered to estimate the muscle force. For long-term evaluations of repetitive movements or static activity, it is interesting to see the power spectrum, knowing that fatigue results in a reduction of the high-frequency signal content. Modern EMG

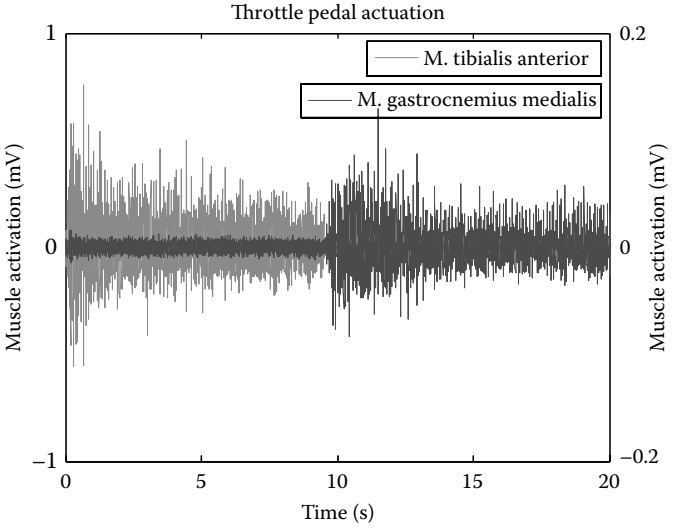


FIGURE 36.5 EMG raw signal of an accelerator pedal actuation. After 10 s (in this case), the antagonist muscle takes over. (Plot generated by Ford Research and Advanced Engineering, Cologne, Germany.)

devices work with 8 or more channels, but the test engineer should keep in mind that too many cables will affect the movement of the test subjects.

36.3.2.5 Climate

Conventional measurement techniques are used to record temperature, moisture, radiation, and air speed. Climate dummies measure the heat transfer at specified body regions with temperature-controlled segments. Unfortunately, there is no common industry standard, and the climate dummies do not work well in hot conditions.

Thermal seat comfort mainly depends on the water vapor permeability because evaporation of sweat is the dominant cooling mechanism of the human body [34]. Component test according to ISO 11092 or DIN EN 31092 can be used to optimize seat covers and reduce the number of climate chamber tests with test subjects. The component test uses a so-called skin model to generate a controlled water evaporation. The water vapor is then guided through the test sample. A high permeability correlates to a high evaporation rate that can be measured quite simply by the power consumption of the evaporation device (Figure 36.6).

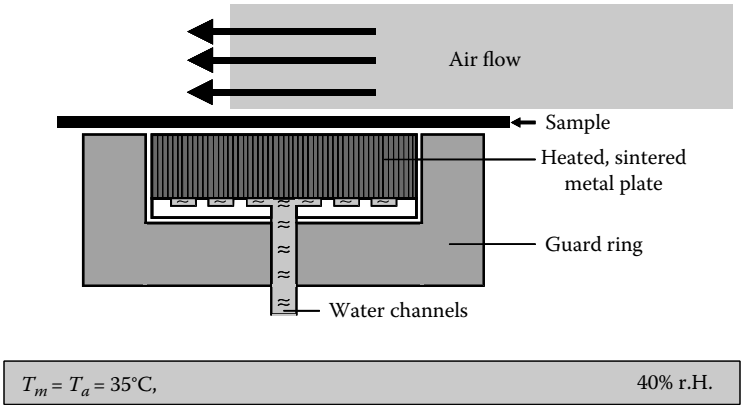


FIGURE 36.6 Skin model according to ISO 11092. (Sketch generated by Ford Research and Advanced Engineering, Cologne, Germany.)

36.3.2.6 Others

Discomfort and fatigue affect several body reactions. Some of the body reactions can be measured quite easily, for instance, the heart rate or the electrical conductivity of the skin. Eye motion tracking, including the blink frequency, is another example.

36.3.3 CAE Models

Compared to other automotive applications, CAE-driven comfort prediction is a very young branch with a low level of standardization throughout the industry. Due to the different comfort aspects and the complexity of the human body, it is not possible to include the entire comfort prediction into one single model. There will always be dedicated models for posture analysis, vibration analysis, or analysis of muscle loads, to name a few. A complete documentation of all proposed modeling strategies would fill at least one book. Some examples will be shown here, without claiming that this is a complete overview.

36.3.3.1 Kinematic Models

Kinematic models such as Ramses [26] use a segmented rigid body representation and determine the joint angles for a given posture. Based on experimental data, a comfort index can be derived from the joint angles. Design optimization in this case is equivalent to optimization of this comfort index. The method can be applied for static seating positions and movement sequences, for example, ingress/egress or pedal operation [36]. Recent studies include a strength model with "torque potatoes" [27].

36.3.3.2 Inverse Dynamic Models without Muscles

Rigid body models using inverse dynamics deliver the required joint torques and forces to fulfill a predefined motion task, taking all external loads into account [39]. Joint reactions have no direct physiological meaning since the human body does not work like a robot, but obviously there is a strong correlation between muscular effort and joint torques. These models are valuable for comparisons and trend predictions. The downside of the inverse dynamics approach is the large amount of required input parameters, which are sometimes difficult to obtain.

36.3.3.3 Inverse Dynamics Models Including Muscles

Adding the muscles to inverse dynamic models means that the modeling effort increases by about one magnitude. For a full-body simulation, several hundred muscles need to be taken into account, each of them with a specified line of action and strength properties [21]. Optimization criteria need to be defined as well because the number of unknowns exceeds the number of equations [29]. The main advantage of this fully developed model is the high bio-fidelity, meaning that the calculated results have a real physiological meaning. Compared to models without muscles, the required test data are the same, assuming a segmented rigid body approach (Figure 36.7).

36.3.3.4 Forward Dynamic Models to Simulate the Human Body Impedance

These models are more or less a virtual vibration dummy and can be used for seat development [5]. A good representation of the human body with respect to the vibration impedance allows the prediction of the seat vibration transmissibility. Obviously, the seat needs to be modeled as well, typically in a functional description with nonlinear springs and dampers. Forward dynamics means that the movement is calculated as a result of all acting forces. The main drawback of these models is the description of the mechanical coupling between seat and occupant. Typically the model focuses on vertical vibrations only.



FIGURE 36.7 AnyBody car driver model (Ford). (Plot generated by Ford Research and Advanced Engineering, Cologne, Germany.)

36.3.3.5 Forward Dynamic Models to Simulate the Vibration of the Human Body

These models are used to calculate the propagation of the seat vibrations into different parts of the human body [7,35]. This is important for further definitions of critical vibration doses or to investigate how people can operate vehicles under vibration exposure. The main advantage is a real analysis of the dynamic behavior in the human body, giving additional information over pure test data. However, these models are highly complex and require advanced modeling of the deformable body parts as well as a representation of the muscular tension, including all anatomical properties.

36.3.3.6 FEM Prediction of the Seat Pressure Distribution

These models focus on the seat pressure distribution under static load. The main task is the proper representation of the seat–occupant coupling including the foam and deformable body tissue [14,19].

36.4 CONCLUSION

When dealing with vehicle comfort, it is very easy to get lost in the exploration of fragmented details. Priorities need to be set during the vehicle development process in order to cover all the aspects of comfort. The future goal will be to replace subjective ratings with reliable test methods or with CAE methods that do not require hardware. Limited resources might force the development teams to accept simplifications when they combine the results coming from different approaches. Knowing that the customer perception of comfort is nothing less than a total score of many independent performance criteria, a good overview and the clever synthesis of the fragments will be crucial. For the engineer, this requires an open mind since the machine is just one part of the man–machine system. Even psychology comes into play as perceptions of different senses need to be orchestrated to a pleasant comfort symphony.

ACKNOWLEDGMENTS

I wish to thank my family for patience and support as well as my colleagues at FFA for proofreading and the donation of figures.

REFERENCES

1. Bitter, T., F. Fritzsche, J. Hartung: Darstellung des Schwingungsverhaltens von Fahrzeuginsassen. Forschungsvereinigung Automobiltechnik Report 189, Frankfurt, Germany, 2005.
2. Brand, P., W.C. Coleman: The diabetic foot. In: Rifkin, H. and D. Porte (eds.): *Diabetes Mellitus*. 4th edn., New York: Elsevier, 1990.
3. British Department of Health: The prevalence of back pain in Great Britain in 1998. Statistical Bulletin 1999/18, London, U.K., 1999.
4. Bubb, H.: Ergonomie und Sitzgestaltung. In: Wilke, H.J. (ed.): *Ergomechanics*. Shaker Publ., Aachen, Germany, 2004.
5. Liang, C.Y. et al.: A biodynamic model for the assessment of human operator performance under vibration environment. SAE paper No. 2005-01-2742, Warrendale, PA, 2005.
6. De Luca, C.J.: The use of surface electromyography in biomechanics, *Journal of Applied Biomechanics*, 13(2), 135–163, 1997.
7. De Rochefort, E. et al.: Detailed modelling of the lumbar spine for investigation of low back pain. SAE paper No. 2005-01-2716, Warrendale, PA, 2005.
8. European Communities: Directive 2002/44/EC. *Official Journal of the European Communities*, L177/13, 2002.
9. Flügel, B. et al.: *Anthropologischer Atlas, Grundlagen und Daten*. Frankfurt, Germany: Edition Wötzel, 1986.
10. Griffin, M.: *Handbook of Human Vibration*. London, U.K.: Academic Press, 1990.
11. Grunwald, M., L. Beyer: *Der bewegte Sinn – Grundlagen und Anwendungen zur haptischen Wahrnehmung*. Basel, Switzerland: Birkhäuser-Verlag, 2001.
12. Kawachi, K. et al.: Visualization and classification of strategy for entering car. SAE paper No. 2005-01-2683, Warrendale, PA, 2005.
13. Kroemer, K., E. Grandjean: *Fitting the Task to the Human*. Philadelphia, PA: Taylor & Francis 1997.
14. Marx, B. et al.: Virtual assessment of seating comfort with human models. SAE paper No. 2005-01-2678, Warrendale, PA, 2005.
15. McMahon, T.: *Muscles, Reflexes, and Locomotion*. Princeton, NJ: Princeton University Press 1984.
16. Mergl, C. et al.: Predicting long term riding comfort in cars by contact forces between human and seat. SAE paper No. 2005-01-2690, Warrendale, PA, 2005.
17. *Merriam-Webster*: Comfort. http://www.merriam-webster.com/dictionary/comfort, online dictionary, 2005.
18. Mitschke, M.: *Dynamik der Kraftfahrzeuge*. 2nd edn., Berlin, Germany: Springer-Verlag, 1990.
19. Montmayeur, N. et al.: Experimental and numerical analyses of seating pressure distribution patterns. SAE paper No 2005-01-2703, Warrendale, PA, 2005.
20. Pauwels, F.: *Gesammelte Abhandlungen zur funktionellen Anatomie des Bewegungsapparates*. Berlin, Germany: Springer-Verlag, 1965.
21. Rasmussen, J., S. Christensen: Musculoskeletal modeling of egress with the anybody modeling system. SAE paper No. 2005-01-2721, Warrendale, PA, 2005.
22. Salvendy, G.: *Handbook of Human Factors and Ergonomics*. New York: John Wiley & Sons, 1997.
23. Schmidt, R.F., H.G. Schaible: *Neuro- und Sinnesphysiologie*. Berlin, Germany: Springer-Verlag, 2005.
24. Schmidtke, H.: *Handbuch der Ergonomie*. Bundesamt f. Wehrtechnik und Beschaffung, Koblenz, Germany, 1989.
25. Schünke, M., E. Schulte, U. Schumacher: PROMETHEUS, Lernatlas der Anatomie. Allgemeine Anatomie und Bewegungssystem. Thieme-Verlag, Stuttgart, Germany, 2004.
26. Seidl, A.: The man model RAMSIS: Analysis, synthesis and simulation of 3-dimensional human body postures, Dissertation, Technische Universität München, München, Germany, 1994.
27. Seitz, T. et al.: FOCOPP – An approach for a human posture prediction model using internal/external forces and discomfort. SAE paper 2005-01-2694, Warrendale, PA, 2005.
28. Shell: *Shell Pkw-Szenarien bis 2030, Flexibilität bestimmt Motorisierung*. Hamburg, Germany: Shell Deutschland Oil, 2004.
29. Siebertz, K.: *Biomechanische Belastungsanalysen unter Berücksichtigung der Leichtbauweise des Bewegungsapparates*. Göttingen, Germany: Cuvillier-Verlag, 1994.
30. Siebertz, K., B. Krabbe: Die Druckverteilungsmessung und ihre klinische Anwendung. Verdonk, A., M. Wieg (Hrsg.): *Biomechanische Vervahren und ihre praktische Anwendung in Diagnostik und Therapie*, Luedenscheid 1994, S. 121–131.

31. Siebertz, K., T. Moerke, D. DeVogel: The strain gauge goniometer, a new sensor to measure dummy joint angles under crash conditions. *SAE World Congress 2000*. In: *Safety Test Methodologies*. SAE 2000 Special Publications, Warrendale, PA, 2000.
32. Silbernagl, S., A. Despopoulos: *Taschenatlas der Physiologie*. Stuttgart, Germany: Thieme-Verlag, 2003.
33. Stubbs, J. et al.: Quantitative method for determining cushion comfort. SAE paper No. 2005-01-1005, Warrendale, PA, 2005.
34. Temming, J.: Fahrzeugklimatisierung und Verkehrssicherheit. Forschungsvereinigung Automobiltechnik Report 177, Frankfurt, Germany, 2003.
35. Verver, M.: Numerical tools for comfort analyses of automotive seating. PhD thesis, TU Delft, Delft, the Netherlands, 2004.
36. Wang, X. et al.: Validation of a model-based motion reconstruction method developed in the REALMAN project. SAE paper No 2005-01-2743, Warrendale, PA, 2005.
37. Wikipedia: Komfort. http://de.wikipedia.org/wiki/Komfort, online dictionary, 2005.
38. Zatsiorsky, V.: *Kinematics of Human Motion*. Champaign, IL: Human Kinetics, 1998.
39. Zatsiorsky, V.: *Kinetics of Human Motion*. Champaign, IL: Human Kinetics, 2002.
40. Zhang, Y., Helander, M.G., Drury, C.G.: Identifying factors of comfort and discomfort in sitting. *Human Factors* 38(3), 377–389, 1996.

37 Subjective and Objective Evaluations of Car Handling and Ride

Gwanghun Gim

CONTENTS

37.1 Introduction 1397
37.2 Subjective Assessment 1398
37.2.1 Straightability 1401
37.2.1.1 Residual Pull 1404
37.2.1.2 Running Straight 1404
37.2.1.3 Torque Steer 1404
37.2.1.4 Braking Straight 1405
37.2.2 Steerability 1405
37.2.2.1 On-Center Feel 1405
37.2.2.2 Steering Response 1407
37.2.2.3 Linearity 1407
37.2.2.4 Steering Precision 1407
37.2.2.5 Steering Angle 1408
37.2.2.6 Steering Addition 1409
37.2.2.7 Steering Effort 1409
37.2.2.8 Returnability 1409
37.2.3 Controllability 1410
37.2.3.1 Balance FA/RA 1410
37.2.3.2 Power-Off Reaction 1411
37.2.3.3 Grip Limit 1412
37.2.3.4 Controllability at Limit 1413
37.2.3.5 Driving Power 1413
37.2.4 Stability 1413
37.2.4.1 Yaw Stability 1413
37.2.4.2 Rear Axle Stability 1414
37.2.4.3 Roll Stability 1414
37.2.4.4 Braking Stability 1414
37.2.4.5 Rollover Stability 1414
37.2.4.6 Bump Steer 1415
37.2.5 Comfortability 1415
37.2.5.1 Primary Ride 1415
37.2.5.2 Secondary Ride 1415
37.2.5.3 Impact Ride 1416

37.3 Subjective Test 1416
37.3.1 Straight-Ahead Maneuver 1416
37.3.1.1 Free Steering in Straight Ahead 1419
37.3.1.2 Fixed Steering in Straight Ahead 1419
37.3.1.3 Powering On/Off in Straight Ahead 1419
37.3.1.4 Braking in Straight Ahead 1419
37.3.2 Steering Maneuver 1420
37.3.2.1 On-Center Steering 1420
37.3.2.2 Off-Center Steering 1420
37.3.2.3 Progressive Steering 1420
37.3.2.4 Limit-Grip Steering 1421
37.3.2.5 Flick Steering 1421
37.3.3 Lane-Changing Maneuver 1421
37.3.3.1 Lane Changing 1421
37.3.3.2 Severe Lane Changing 1423
37.3.4 Cornering Maneuver 1424
37.3.4.1 Circuit Driving 1426
37.3.4.2 Powering On/Off in a Turn 1430
37.3.4.3 Braking in a Turn 1430
37.3.4.4 Rollover Cornering 1430
37.3.5 Riding Maneuver 1432
37.3.5.1 Wavy Road Driving 1432
37.3.5.2 Rough Road Driving 1432
37.3.5.3 Impact Object Driving 1433
37.4 Objective Test 1433
37.4.1 Handling Test 1433
37.4.1.1 Steering Wheel Sensor 1433
37.4.1.2 Speed Sensor 1434
37.4.1.3 Acceleration Sensor 1435
37.4.1.4 Gyro Sensor 1437
37.4.1.5 GPS 1437
37.4.2 Ride Test 1437
37.4.2.1 Front Wheel 1439
37.4.2.2 Rear Wheel 1439
37.4.2.3 Vehicle Floor 1439
37.4.2.4 Seat Rail 1439
37.4.2.5 Seat Cushion 1441
37.4.2.6 Seat Back 1441
37.5 Vehicle Dynamics Variables 1443
37.5.1 Vehicle Motion Variables 1443
37.5.2 Physical Perception Variables 1447
37.5.3 Objective Measurement Variables 1452
37.6 Vehicle Dynamics Analysis 1453
37.6.1 Formulation 1453
37.6.1.1 Steer Gradient 1454
37.6.1.2 Understeer Coefficient 1454
37.6.1.3 Oversteer Coefficient 1456
37.6.1.4 Power-Off Reaction Steer Coefficient 1456

37.6.2 Analysis 1457
37.6.2.1 Quasi Steady-State 1459
37.6.2.2 Lane Changing 1459
37.6.2.3 Severe Lane Changing 1461
37.6.2.4 Circuit Driving 1464
37.6.2.5 Powering On/Off in a Turn 1464
37.6.3 Conclusion 1467
Acknowledgment 1474
References 1474

37.1 INTRODUCTION

A conceptual target of desirable vehicle performances can be (1) *enjoyable in driving*, (2) *clean feel to steer*, (3) *easy to control*, and (4) *safe to maneuver*. This concept is substantially based on (a) *comfort*, (b) *agility*, (c) *precision*, and (d) *safety*. As a category of vehicle dynamics, the performances are specifically represented by *handling* and *ride*. Nowadays, they are finally evaluated only by a professional driver's subjective assessment through numerous tests, even though test and assessment methods are varied dependent on each automotive or tire company. Even within the same company, the methods may be different depending on each test engineer.

Subjective assessment has traditionally been carried out by specially trained engineers or technicians but not vehicle dynamists. Since in general they do not have extensive vehicle dynamics knowledge, a fundamental vehicle dynamics concept on the subjective assessment has not been fully established. In order to successfully characterize the performances, this concept should be fundamentally understood. It should be noted that *without his or her own experience in subjective assessment, anyone does hardly understand other's subjective assessment*.

Each test engineer has his or her own subjective test and assessment techniques that have been built through various practices done by expert trainers, seniors, and himself or herself. In reality, most test engineers just know how to make subjective rating on the performances. However, they cannot express their perceptual feeling using a scientific language. That is why there has always been an untouchable barrier in the development of vehicle dynamics. During the last half century, vehicle dynamics has been developed to play a role for better handling and ride performances [1–10]. Compared to other technologies, however, the vehicle dynamics has been contributing not much.

For subjective tests to assess handling and ride performances, any practical documentation is seldom founded in public. Fortunately, ISO international standard provides very well standardized procedures for various handling objective tests [11–17]. However, ISO international standard or any other documentation for practical ride objective tests is not available in public yet. Instead, ISO international standard 2631-1 presents measurement methods for periodic, random, and transient whole-body vibration affecting health, comfort, and motion sickness [18].

Some studies provided numerous correlation approaches between objective measurement and overall handling perception [19–21]. Crolla et al. performed a comprehensive study to make handling assessment using a combined subjective–objective approach [22]. Norman and Farrer separately carried out practical approaches for the objective evaluation of on-center handling performance [23,24]. Salaani et al. did a nice study for the experimental appraisal of steering feel for on-center and off-center driving [25]. Some other works tried to practically correlate between the objective measurement and the subjective assessment of transient steer characteristics [26–28].

Numerous studies objectively evaluated ride quality using human discomfort perception [29–35]. Ushijima et al. practically studied a correlation between the objective measurement and the subjective assessment of impact harshness [36]. Some other works also carried out for the objective evaluation of the impact harshness using only vibration effect [37,38]. Amman et al. considered sound and vibration contributions to the human perception of the impact harshness [39].

When we are talking about handling performance, several words are alternatively used such as straight line, steering, cornering, and stability [40–43]. The handling performance is believed as the combined characteristic of a vehicle and a driver. In reality, its subjective assessment is generally affected by a driver's assessment capability and preference. It is very important that this performance is assessed with the least influence of the driver himself.

The subjective assessment and subjective test given in this handbook are very practical ones commonly used in a homologation process of a car or a tire by many automotive and tire companies worldwide. Here, the subjective assessment and subjective test are tried to be standardized based on their physical concepts. As a tool to make the subjective assessment, the subjective test is described. For each subjective test, its basic concept, test procedure, and driving condition are explained. Even if its concept is agreed with that of other test methods, its procedure and/or condition may be possibly different from the others.

For the purpose of well organizing the subjective assessment, the handling and ride performances can be primarily grouped in five *capabilities*. Each capability is categorized into correspondent *characteristics*. Similar to the subjective assessment, the subjective test is principally classified in five *maneuvers*. Each maneuver is composed of several *tests* with specially designed procedures.

As a supplemental method to the subjective test, objective test is presented to obtain objective measurement data eventually used for a vehicle dynamics analysis. Using vehicle dynamics variables extracted from the measurement data, the vehicle dynamics analysis is discussed to explain the subjective assessment. For that purpose, (a) *subjective assessment variables*, (b) *objective measurement variables*, and (c) *physical perception variables* are defined here.

In order to subjectively assess each characteristic of handling and ride performances, the subjective assessment variables are perceptually evaluated. The objective measurement variables are employed in order to objectively describe vehicle motion representing the characteristic. The physical perception variables are physical variables that test engineers actually perceive. Eventually, these variables can be expressed in terms of the objective measurement variables.

37.2 SUBJECTIVE ASSESSMENT

SAE J1441 explains a 10-point subjective rating scale for vehicle handling [44]. However, it does not give how to make subjective assessment. In practice, most of automakers and tire companies are using this scale.

In this handbook as shown in Figure 37.1, the handling and ride performances of a vehicle are grouped by five *capabilities* such as (1) *straightability*, (2) *steerability*, (3) *controllability*, (4) *stability*, and (5) *comfortability*. Each capability is specifically categorized into several *characteristics*. The nomenclatures of the characteristics given in Figure 37.1 have been traditionally employed to express their perceptual concepts but not their scientific definitions [40–43]. Their names may be not suitable in engineering sense and also a little different depending on each user. All characteristics are supposedly independent from each other, while they are cross affected by their causes. Even different characteristics may share the same cause.

The straightability is a straight-running ability without any significant lateral deviation from the straight path regardless of free or fixed steering. The steerability is a steering capability to achieve intended steering reactions with preferred effort and better comfort of steering. During any vehicle maneuver, easiness to control a vehicle in following an intended path is the controllability. The stability is the convergence characteristic of vehicle behavior even beyond grip limit. The comfortability is the ride comfort characteristic of a vehicle running over uncomfortable, uneven, or irregular roads.

The straightability is primarily affected based on the reaction of front tires due to external disturbances either with free or fixed steering. The response of the front tires against a series of steering applications and the following-up capability of the rear tires against the behavior of the front tires determine the steerability, controllability, and stability. The roll and pitch motions of a sprung mass are also critical factors to affect the stability as well as the controllability.

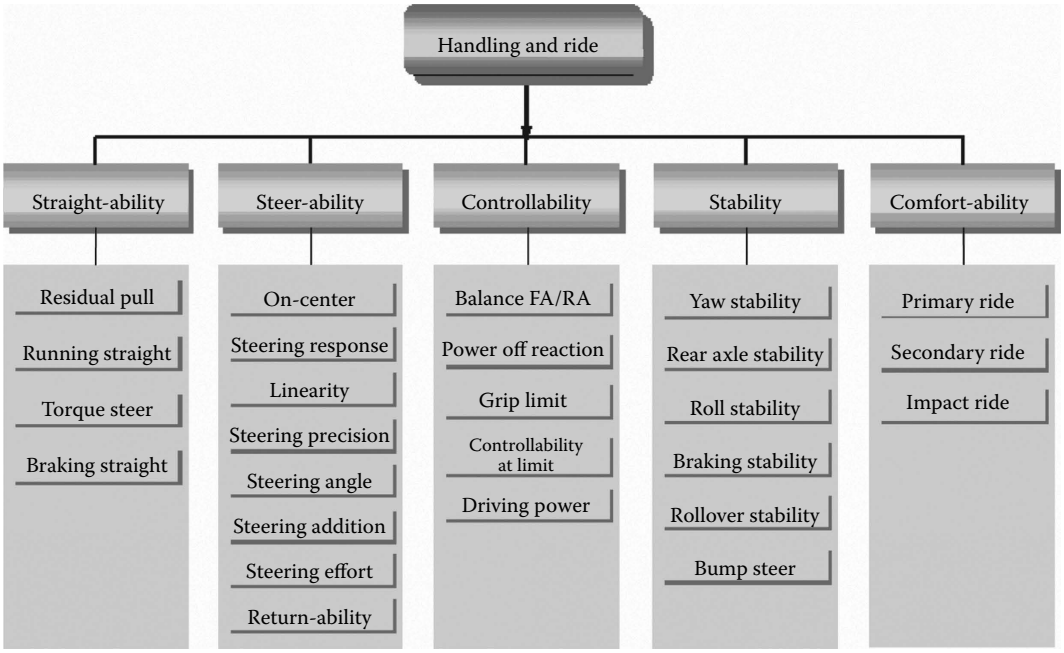


FIGURE 37.1 Subjective assessment on handling and ride performances.

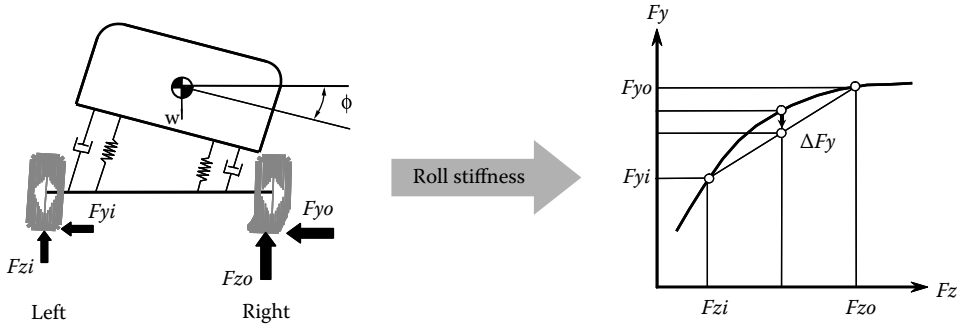


FIGURE 37.2 Lateral load transfer from inner tires to outer tires.

Roll resistance produces a load transfer from inner tires to outer tires while cornering, and then it reduces total cornering force at each axle (refer to Figure 37.2). This reduction of the cornering force not only causes grip reduction but also changes the grip balance between front and rear axles. This balance principally determines the steer characteristic of a vehicle such as transient understeer and oversteer. In case it causes too much understeer or/and oversteer, it deteriorates handling performance due to poor controllability and even bad stability.

While powering on in straight ahead as shown in Figure 37.3a, a front axle loses some grip due to a front-to-rear load transfer. On the other hand, a rear axle gains more grip. With a front-wheel-driving vehicle running on low-friction road, powering-on operation deteriorates even more front axle grip based on a so-called friction ellipse concept. In the situation of powering on in a turn (refer to Figure 37.3b), a vehicle behaviors much more understeer, which makes a driver more difficult to control the vehicle to follow an intended path.

During powering off or braking in straight ahead, a rear-to-front load transfer reduces rear axle grip, while front axle grip gains more as shown in Figure 37.4a. In a turn, a rear-wheel-driving

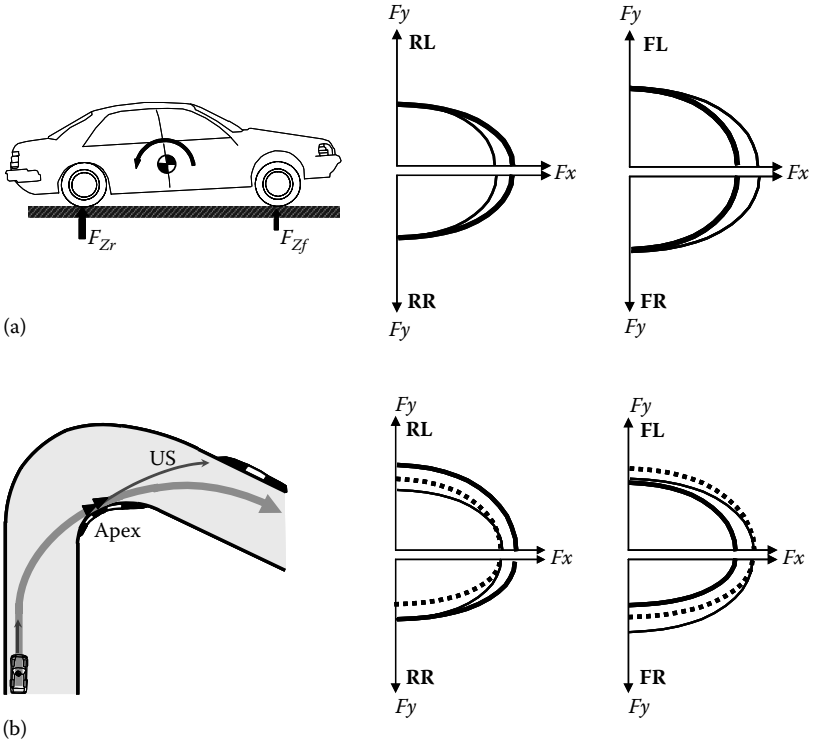


FIGURE 37.3 Longitudinal load transfer and tire force changes due to (a) power on in straight ahead and (b) power on in a turn.

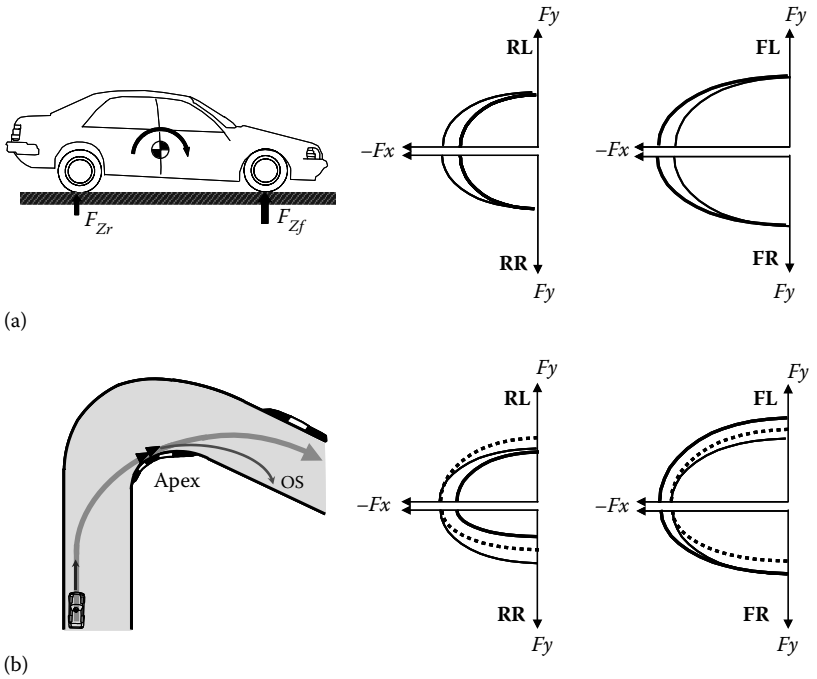


FIGURE 37.4 Longitudinal load transfer and tire force changes due to (a) power off or braking in straight ahead and (b) power off or braking in a turn.

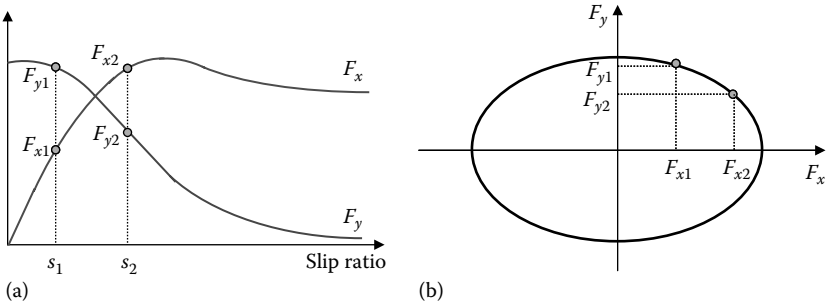


FIGURE 37.5 Friction ellipse concept. (a) Lateral force against longitudinal force and (b) friction ellipse.

vehicle running on low-friction road deteriorates even more rear axle grip based on the friction ellipse concept (refer to Figure 37.4b). Sometimes it produces big oversteer beyond control limit in a turn so that it is easy for a normal driver to get into an accident.

An application of power on, power off, or braking operation to a tire having lateral slip in cornering increases longitudinal force in magnitude. Instead, it reduces lateral force in magnitude due to the generation of longitudinal slip in addition to the lateral slip as shown in Figure 37.5a. The maximum potential of the resultant tire friction force is observed to be limited by about an ellipse (refer to Figure 37.5b). This phenomenon is called the friction ellipse concept. Figures 37.6 and 37.7 illustrate the friction ellipses of all tire forces of a vehicle in a right turn on high-friction road (dry surface) and low-friction road (wet surface), respectively.

When test engineers carry out their subjective assessment, they make their judgment using *subjective assessment variables*. Through special trainings, self-trainings, and test experiences, up to nowadays, test engineers learn these variables just in forms of feeling but not in physical forms. The test engineers believe that all these variables are tactically, visually, and audibly perceived during various subjective tests. However, they do not directly sense all of these variables. Instead, in reality, they do perceptually measure *physical perception variables*, which are discussed in detail later.

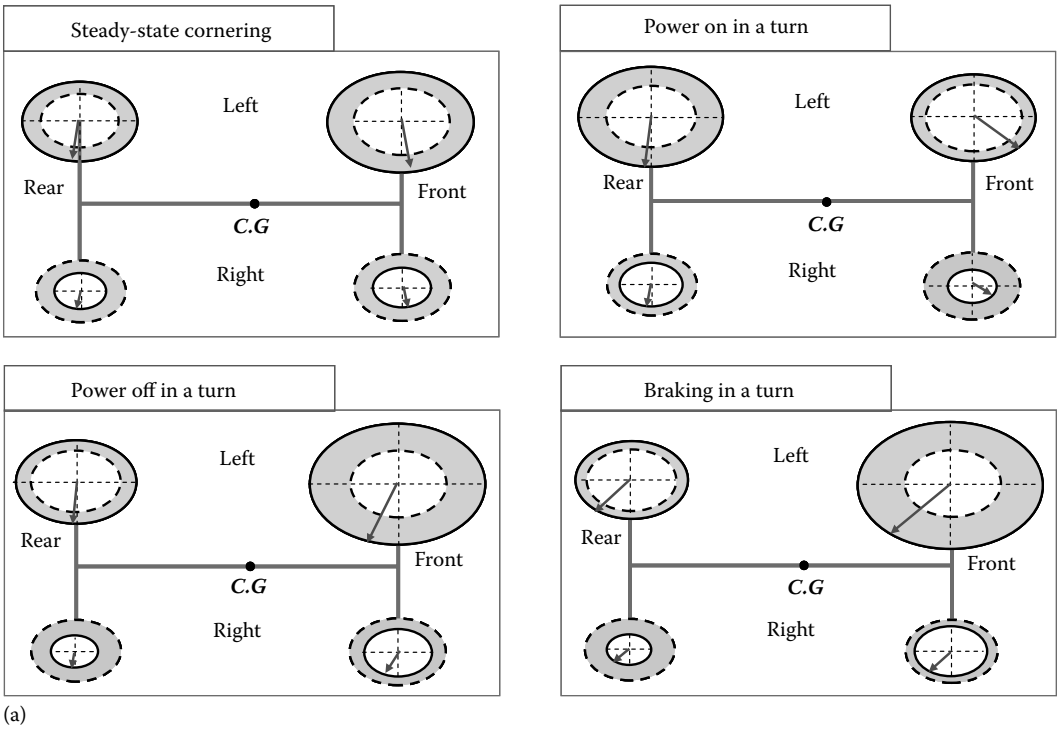
As the subjective assessment variables regarding steering behavior, *steering handwheel angle* is visually measured, and *steering handwheel torque* is tactically sensed. The amount of required steering handwheel angle is used as a guideline for steering gain. The same amount of the steering handwheel angle is applied to evaluate the steering responses of different vehicles. The strength of reacted steering handwheel torque is estimated for steering effort.

As the primary subjective assessment variable on vehicle cornering or lane-changing behavior, *vehicle path* in addition to *vehicle speed* is subjectively evaluated during straight running, cornering, lane changing, or severe lane changing. On top of that, *vehicle attitude* is also subjectively assessed. In this handbook, the vehicle attitude is defined with respect to vehicle moving direction. Therefore, the vehicle attitude stands for *vehicle nose-in attitude relative to vehicle path at an instant*.

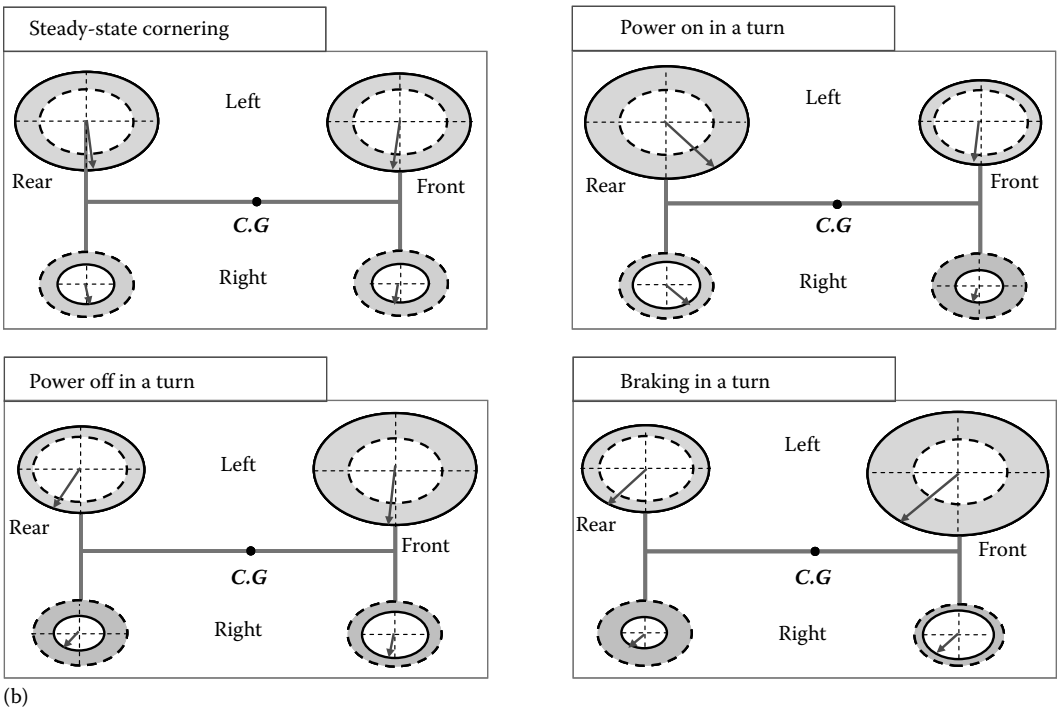
For roll stability and rollover stability, *vehicle body motion* is sensed through both driver body feel and visual observation. In order to subjectively assess comfortability, the vibratory motions of sprung and unsprung masses are tactically sensed. In addition, tire impact shock is tactically and audibly preserved.

37.2.1 Straightability

Regardless of any disturbance in driving condition, straightability to make a vehicle run in straight ahead is more often called straight-line stability. This nomenclature is just a traditionally used perceptual one but not from its physical definition. The straightability usually represents *comfort*, but in

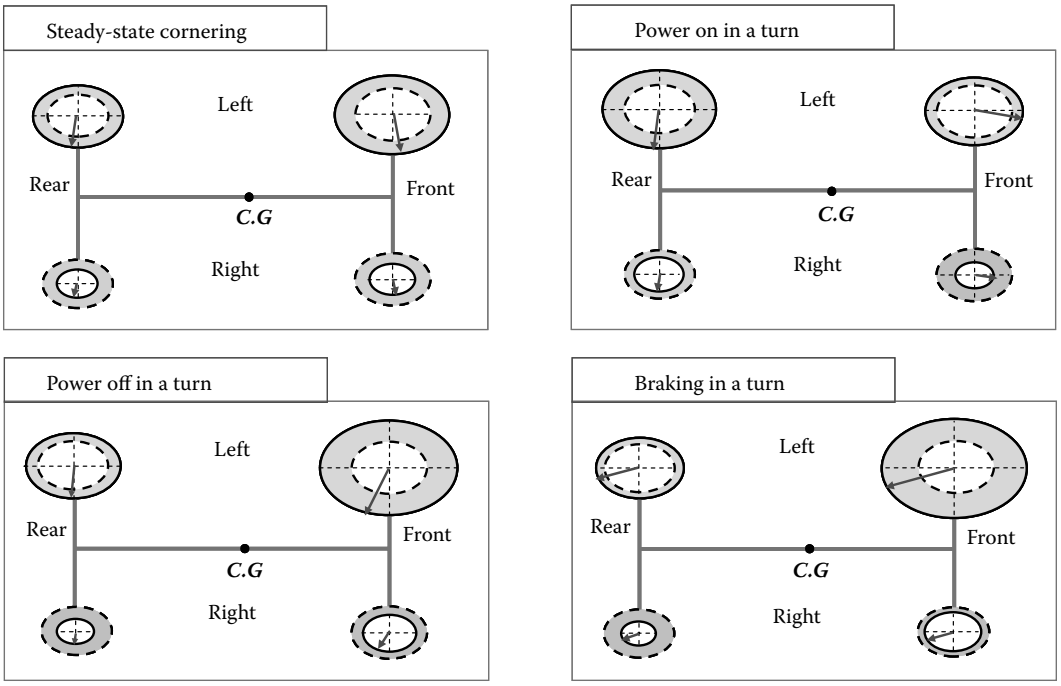


(a)

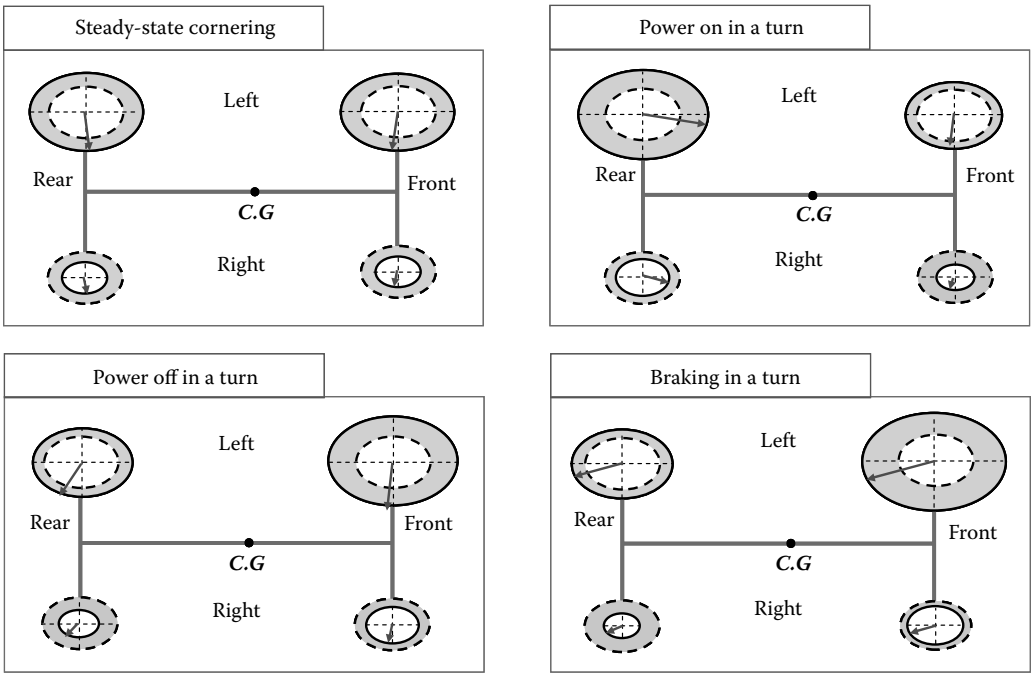


(b)

FIGURE 37.6 Friction ellipses of all tire forces of a vehicle in a right turn on high-friction road (dry surface). (a) Front wheel driving and (b) rear wheel driving.



(a)



(b)

FIGURE 37.7 Friction ellipses of all tire forces of a vehicle in a right turn on low-friction road (wet surface). (a) Front wheel driving and (b) rear wheel driving.

the worst case, it is related to *safety*. In general, it is assessed through the perceptual evaluations of its four characteristics such as (a) *residual pull*, (b) *running straight*, (c) *torque steer*, and (d) *braking straight* (refer to Figure 37.1).

37.2.1.1 Residual Pull

Residual pull is the lateral drift characteristic of a vehicle caused due to residual steering torque. It is subjectively or objectively assessed by the magnitude of lateral drift from straight path after about 100 m driving with free steering at constant speed (usually 100 kph) as shown in Figure 37.8. It confirms the proper design of a steering system, suspension setup and tire matching, and their manufacturing uniformities. While long driving, a residual pull problem makes a driver feel tired due to the residual steering torque. Furthermore, with free steering, it may cause an accident in the worst case.

37.2.1.2 Running Straight

Running straight is a straight-running ability without any drift, wander, or sloppiness with fixed steering at constant speed as shown in Figure 37.9. It illustrates the design sensitivity of steering and suspension systems against road irregularity or wind disturbance. Its problem with long driving also makes a driver feel tired due to continuous steering corrections.

37.2.1.3 Torque Steer

Torque steer is the lateral drift characteristic of a vehicle produced by an unbalance of traction torques between left and right driving wheels, when powering on in straight ahead with fixed steering. It is subjectively evaluated with the lateral drifting motion of the vehicle moving to an adjacent lane with full power on and back to the original lane with power off as shown in Figure 37.10. Smaller lateral drift is desired for better torque steer.

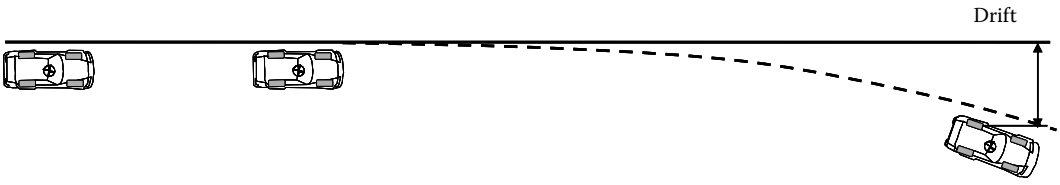


FIGURE 37.8 Residual pull.

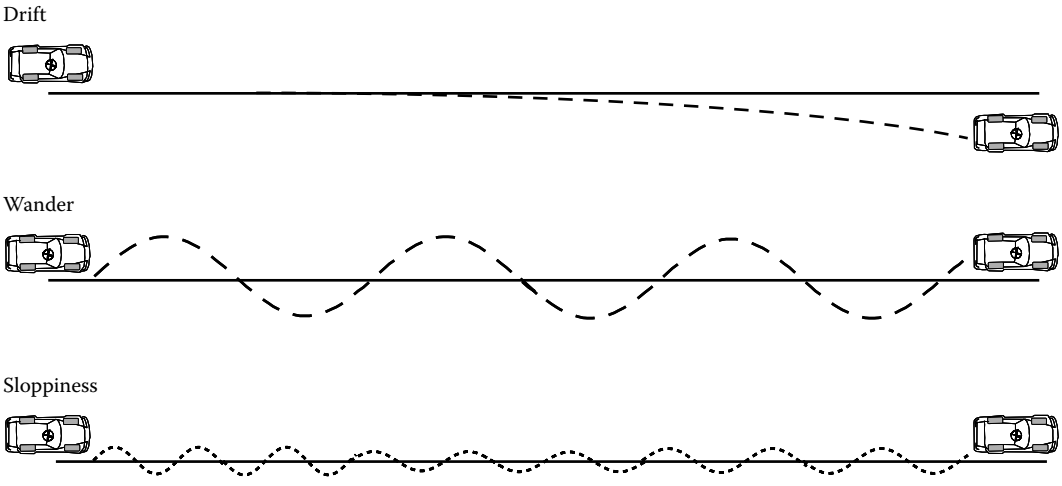


FIGURE 37.9 Running straight.

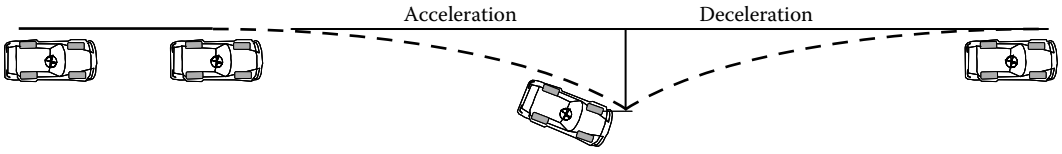


FIGURE 37.10 Torque steer.

37.2.1.4 Braking Straight

Braking straight is a straight-braking ability without any disturbance of vehicle attitude or path, when a vehicle instantaneously stops completely. It is subjectively assessed by the deviation of vehicle attitude or path during full braking with hands-on steering. No deviation of vehicle attitude or path is pursued for good braking straight.

37.2.2 Steerability

The major aims of steering behavior to be achieved are (1) *comfort*, (2) *agility*, and (3) *precision*. The steering behavior is represented by steerability categorized in eight characteristics such as (a) *on-center feel*, (b) *steering response*, (c) *linearity*, (d) *steering precision*, (e) *steering angle*, (f) *steering addition*, (g) *steering effort*, and (h) *returnability* (refer to Figure 37.1). The steerability is subjectively assessed with steering handwheel torque and vehicle path as output in response to an input of steering handwheel angle. Depending on the zone of a steering input, the steerability is specified as shown in Figure 37.11.

On-center feel, linearity, and steering addition are mild transient steering characteristics. On the other hand, all others are transient characteristics.

37.2.2.1 On-Center Feel

On-center feel is a comprehensive perception on (1) *steering torque* and (2) *vehicle path change* in response to a very slow steering input at and around the center of steering. Figure 37.12 presents the on-center steering torque represented by (a) *friction*, (b) *hollow band*, (c) *compliance*, and (d) *steering feedback*. The vehicle path change (or vehicle turn) is perceived against steering handwheel angle to assess (a) *dead band* and (b) *steering gain* as shown in Figure 37.13. The on-center feel must be clear and exact; otherwise, a driver feels nervous. Since the desirable magnitude of the on-center steering torque is dependent on each customer, it is one of the important items determining a vehicle concept.

The steering friction is initial steering torque at the center of steering. In principle, the least steering friction is desirable. However, it is practically not easy to be achieved. The improvement

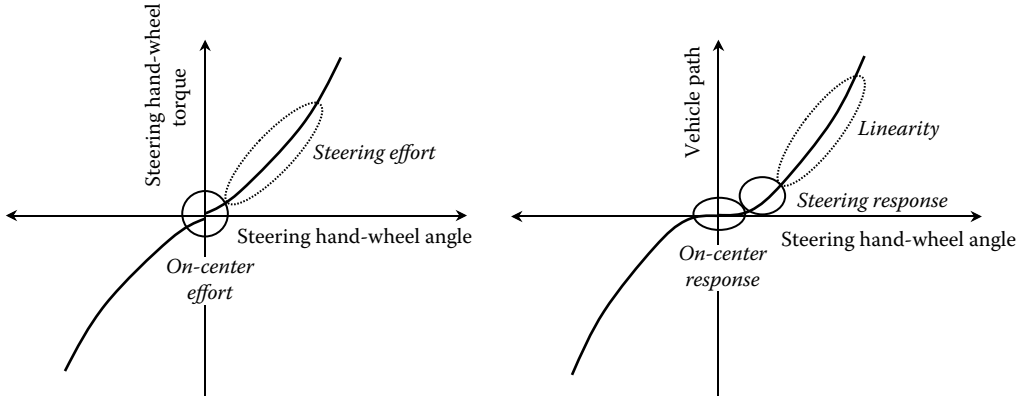


FIGURE 37.11 Steering handwheel torque and vehicle path specifying steerability.

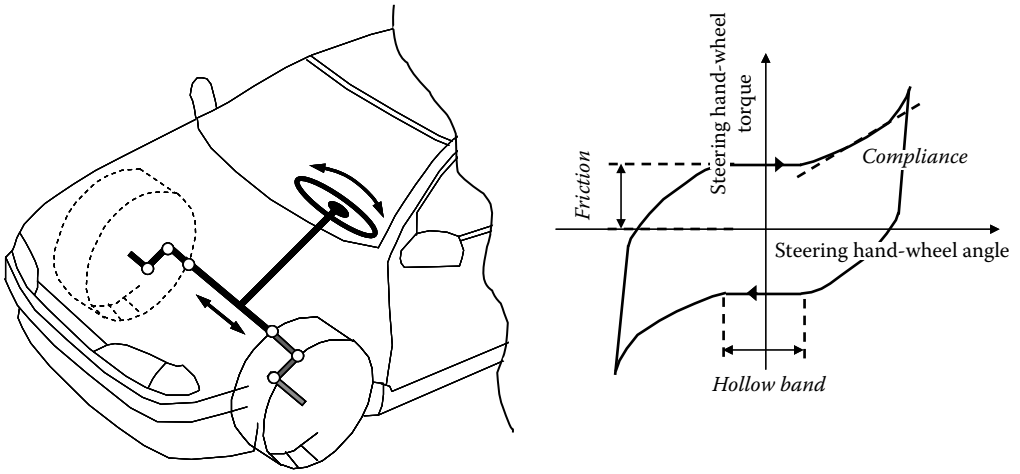


FIGURE 37.12 On-center steering torque.

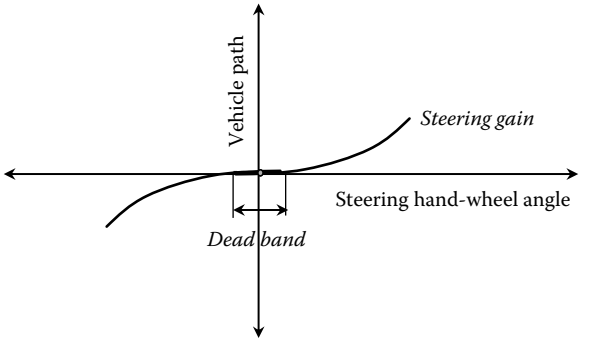


FIGURE 37.13 On-center steering response.

of the steering friction may cause some other drawbacks on steering sensitivity. In the case of big friction, it seriously hurts steering behavior.

With a little steering application from the steering center, the hollow band is sensed as the band of steering looseness caused due to no steering torque generation in addition to the friction. No hollow band is generally pursued. Otherwise, the steering effort is felt just loose around the steering center with a hollow band. In case the hollow band is significantly big, it causes too much time delay of steering response with consequent sudden vehicle path change.

The compliance is assessed with the perception of progressive steering movement against a continuous and smooth incremental steering application from the steering center. The ideal target of this perception is just like the feeling of spring compression. A proper mechanical terminology for this property should be "*stiffness*." However, "*compliance*" may be traditionally used to express the perceptual feeling of the compression.

The dead band is just the band of no vehicle path reaction generated in response to a miniature steering application from the steering center. When no dead band exists, steering corrections are always required corresponding to road irregularity and wind change. It bothers a driver to do so at every instant. On the other hand, a big dead band provokes too much time delay of steering response. It consequently produces rapid vehicle path change, which gets on a driver's nerves.

The steering gain represents a perceptual measure of an output-to-input ratio. It is assessed depending on how big the initial path change of a vehicle is generated from the steering center.

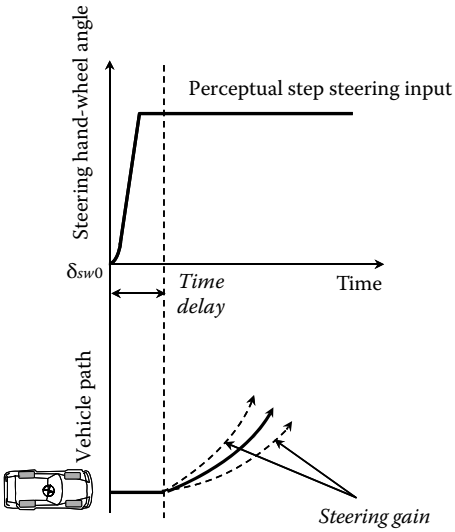


FIGURE 37.14 Off-center steering response.

It requires a proper degree of the sharpness of the vehicle path change (or vehicle turn) depending on a driver's preference, that is, not too big or not too small.

The steering feedback is feedback information transferred to a steering handwheel from tires through a steering system. In the case of good steering feedback characteristics, a driver can feel an interaction between tires and road surface. Then, it helps the driver to predict upcoming road condition. In practice, it is not easy to achieve a great level of the steering feedback characteristic.

37.2.2.2 Steering Response

Steering response is an overall subjective measure of time-dependent vehicle path change against a transient steering input. Figure 37.14 shows the vehicle path versus steering handwheel angle containing (a) *time delay* and (b) *steering gain*. Both time delay and steering gain represent the subjective quality of *agility* of the vehicle path change in time. The steering response is assessed while off-center steering both in straight ahead and in a turn.

The time delay is just a time lag until the first movement of directional change of the vehicle path. Less time delay is usually felt better. The steering gain represents a subjective measure of steering response how fast the initial movement of direction change is developed after its start of generation. It requires a proper degree of the sharpness of the vehicle path change depending on a driver's preference, that is, not too fast or not too slow.

37.2.2.3 Linearity

Linearity is a perceptual impression on the linearity characteristic of trajectory of vehicle path in response to a gradually incremental steering input. Figure 37.15 shows the profile of vehicle path to describe the linear tendency of steering gain. Any change of vehicle attitude is additionally perceived.

Good linearity characteristic requires the linearly progressive change of the vehicle path against the gradually incremental steering input. In addition, it is naturally required constant vehicle attitude. When there exists any disturbance of the vehicle path or attitude, in other words, it is assessed badly.

37.2.2.4 Steering Precision

Steering precision is an overall subjective impression on how accurately a vehicle follows an intended path with a series of steering maneuvers. While path following, the easiness to control the vehicle also affects subjective rating. It is usually evaluated as the characteristic of *precision* or *line*

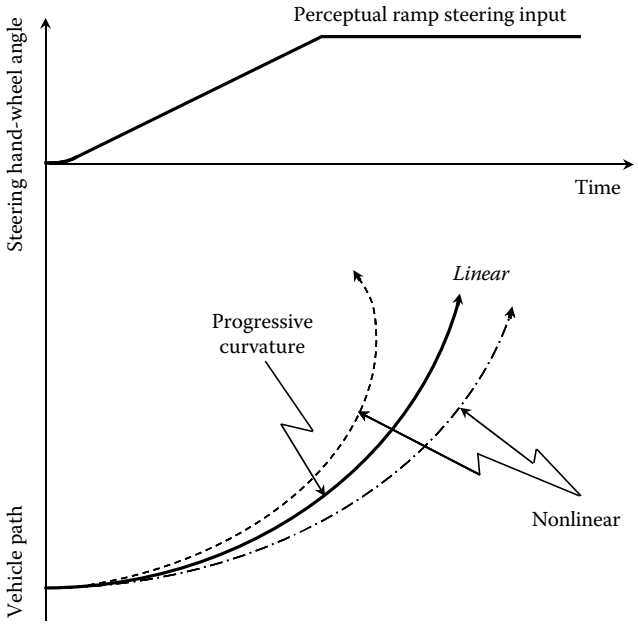


FIGURE 37.15 Linearity.

trace when path following in a handling circuit. In addition, it is judged with the accuracy of lane changing and severe lane changing.

37.2.2.5 Steering Angle

Steering angle is a subjective evaluation of an output-to-input ratio. It is assessed by the amount of steering application to get the same path. In general, the smaller the angle, the better. For example, Figure 37.16 shows the steering gains of a vehicle equipped with two different tire sets. Tire A has faster steering response but needs more steering angle to get the same path change than Tire B. The steering angle is evaluated while off-center steering both in straight ahead and in a turn.

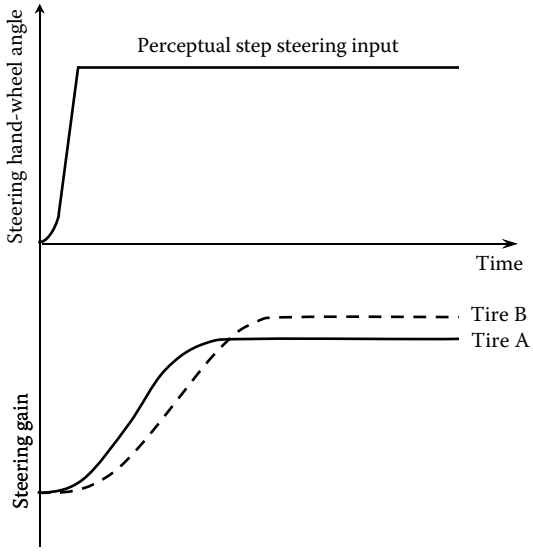


FIGURE 37.16 Steering angle.

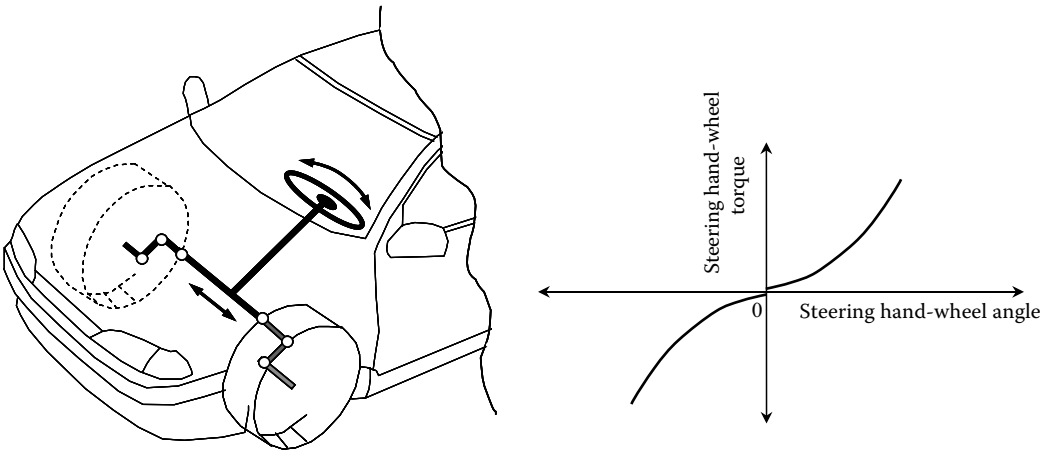


FIGURE 37.17 Steering effort.

37.2.2.6 Steering Addition

Steering addition is a perceptual indicator of additional grip produced when an extra steering input is applied at grip limit in a turn. In this situation, some path change is subjectively assessed. An understeer behaved vehicle usually requires better steering addition characteristic to come back to the intended path in a corner. In general, the more additional grip, the better.

37.2.2.7 Steering Effort

Steering effort is a subjective measure of steering handwheel torque required to change the direction of vehicle path. As shown in Figure 37.17, it should not be too loose or too tight but appropriate for the best *comfort*. The steering effort is assessed through off-center steering both in straight ahead and in a turn.

37.2.2.8 Returnability

Returnability is a perceptual evaluation of quickness and precision of steering handwheel movement returning to its neutral position primarily just after a flick steering input is applied and immediately released (refer to Figure 37.18). It represents the ability of tires to make the steering handwheel come

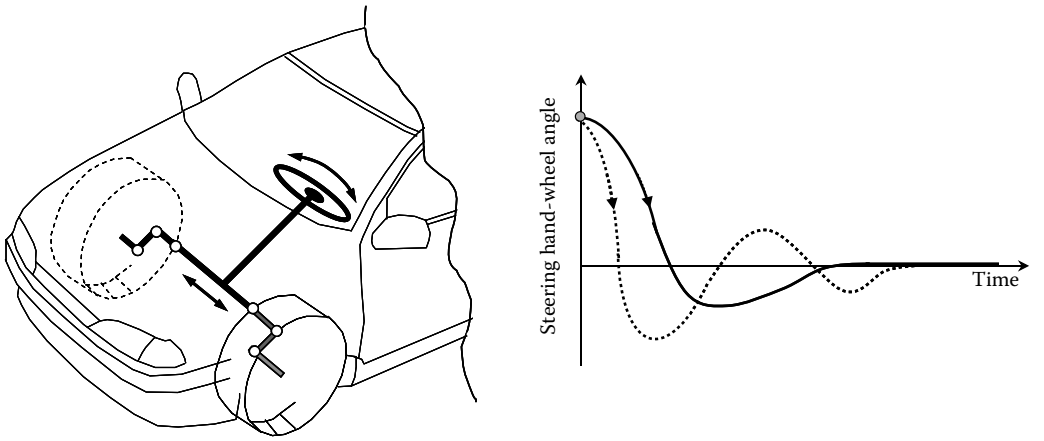


FIGURE 37.18 Returnability.

back directly and smoothly to the steering center after flick steering as well as while exiting a corner. Less oscillation of the steering handwheel and the smoothness of its movement are desirable.

37.2.3 Controllability

Controllability is a psychological assessment of easiness to control a vehicle in following an intended path in a handling circuit or while lane changing. This capability is basically dependent on the grip capabilities of front and rear axles as well as a grip balance between them. The controllability is composed of its five characteristics such as (a) *balance FA/RA*, (b) *power-off reaction*, (c) *grip limit*, (d) *controllability at limit*, and (e) *driving power.*

37.2.3.1 Balance FA/RA

Balance FA/RA is an abbreviation of a *balance* between *front axle grip* and *rear axle grip*. In practice, it stands for a *balance* between *transient understeer* and *transient oversteer* [40–43,45–47]. Through various handling tests, the worst (or maximum) levels of the understeer and oversteer are perceptually evaluated. Then, the balance FA/RA is assessed based on an overall impression on these levels.

As shown in Figures 37.19 and 37.20, the transient understeer and transient oversteer are subjectively assessed in a turn. Both their ideal targets are just *steady-state neutral steer* to exactly follow an ideal

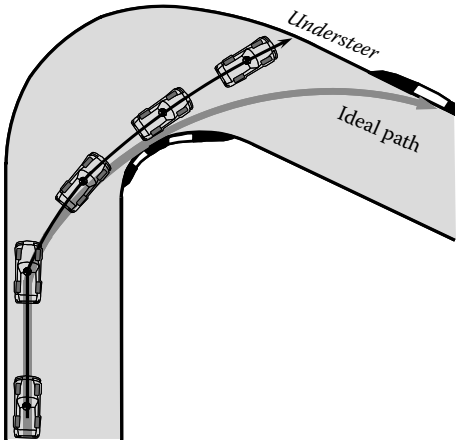


FIGURE 37.19 Transient understeer.

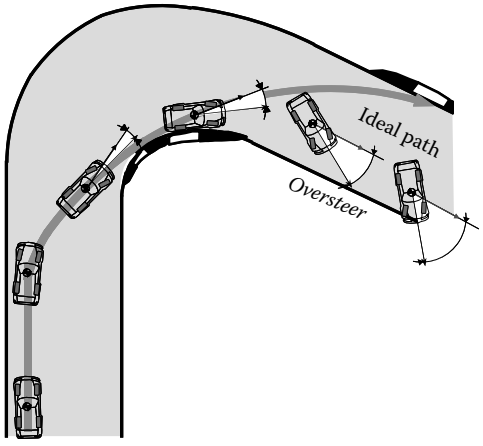


FIGURE 37.20 Transient oversteer.

vehicle path with desirable *vehicle attitude* relatively aligned along the vehicle path in a turn. In other words, the less understeer or oversteer closed to the neutral steer, the better. The vehicle path and attitude are exclusively dependent on the levels of front axle grip and rear axle grip relative to each other.

The transient understeer is a perceptual measurement of front axle grip in a turn. Bigger understeer indicates less front axle grip that provides less capability to precisely follow an ideal vehicle path. It is subjectively assessed as smaller or bigger understeer primarily depending on a smaller or bigger resultant turn, respectively. Each test engineer makes his or her own subjective rating by comparing his or her perceptual vehicle path against his or her own reference value, that is, the vehicle path of the neutral steer.

Similar to the transient understeer, the transient oversteer is a perceptual barometer of rear axle grip in a turn. Bigger oversteer stands for less rear axle grip that induces more slip at a rear axle. Then, it produces outward sway and then bigger vehicle attitude relatively away from vehicle moving direction in a turn. It is subjectively assessed bigger oversteer by estimating driver's perceptual vehicle nose-in attitude relative to his or her own reference value representing the neutral steer. Big oversteer is normally not desirable, because it is more difficult for a normal driver to control a vehicle having a big oversteer tendency.

For example [45], Figure 37.21 shows the transient steer characteristic of a racing car during cornering. It is featured transient understeer and transient oversteer in the zones of the entry and the exit, respectively. In the midcorner where a steady-state condition is usually attained, there exists a transition from the understeer to the oversteer. Since the racing car is equipped with a rear engine and rear driving wheels, it loses some cornering force at a rear axle when it fully accelerates in the exiting zone. That's why this car behaves transient oversteer while exiting in a corner, especially at low-friction surface.

37.2.3.2 Power-Off Reaction

Power-off reaction is a criterion of grip loss of a rear axle caused due to an instantaneous load transfer from the rear axle to a front axle while powering off immediately just after full powering on with fixed steering in a turn (refer to Figure 37.22). The sudden load shift results in an impulsive increase and a simultaneous decrease of cornering forces at the front and rear axles, respectively. The change of cornering forces does quickly increase a yaw moment momentarily producing more vehicle nose-in attitude.

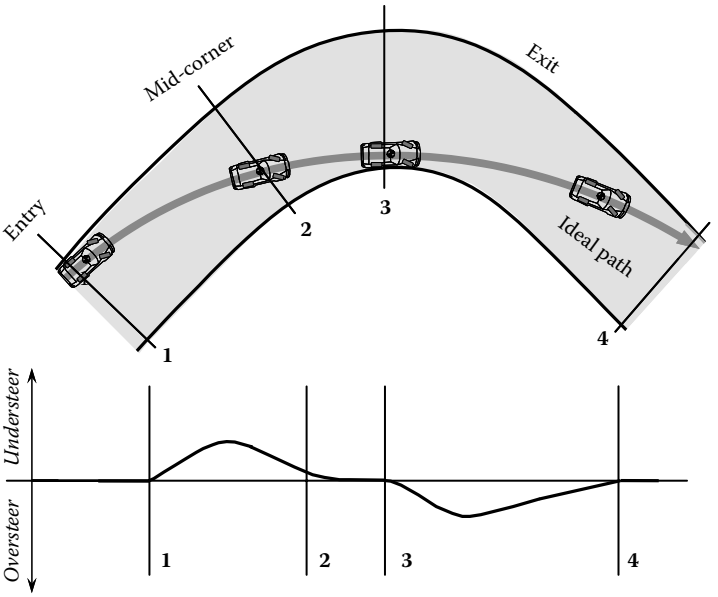


FIGURE 37.21 Transient understeer and oversteer.

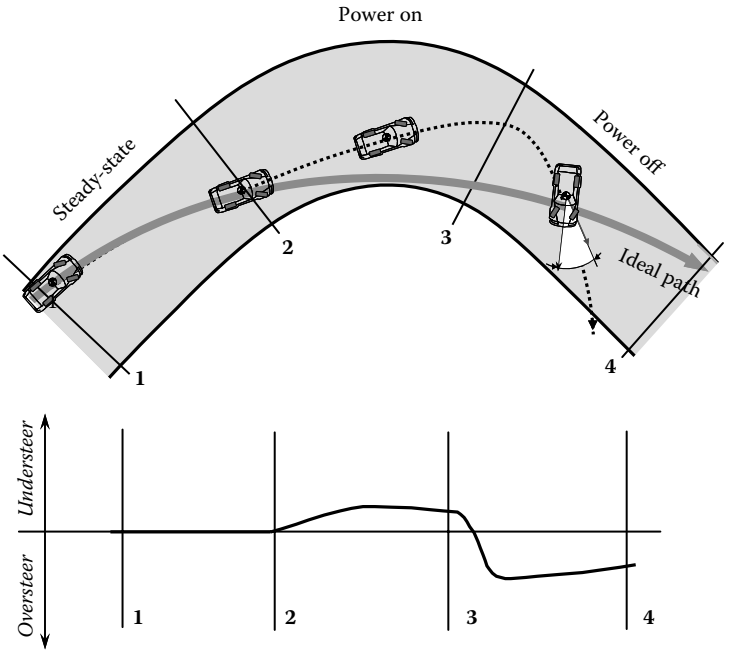


FIGURE 37.22 Power-off reaction steer.

The power-off reaction steer is defined to generally describe the quickness of steer variation from the transient understeer to the transient oversteer as shown in Figure 37.22 [45]. In practice, the power-off reaction steer is subjectively assessed with a driver's perceptual vehicle nose-in attitude change with respect to a time interval from full power on to instantaneous power off. The levels of the understeer and oversteer are additionally evaluated. For better power-off reaction, (a) a slow and smooth power-off reaction steer as well as (b) small transient understeer and (c) small transient oversteer is desirable. It is easy to control a vehicle conducting better power-off reaction behavior.

37.2.3.3 Grip Limit

Grip limit represents the attainable maximum speed in still controlling a vehicle in lane changing or cornering. It is assessed by the speed limit to be achievable at the worst situation. This grip limit is dependent on not only the *grip capabilities* (or cornering forces) of front and rear axles but also their *grip balance*. As shown in Figure 37.23, the cornering forces of the front and rear axles should be more progressive but without their sudden drops beyond the slip angle causing the grip limit for better *predictability*.

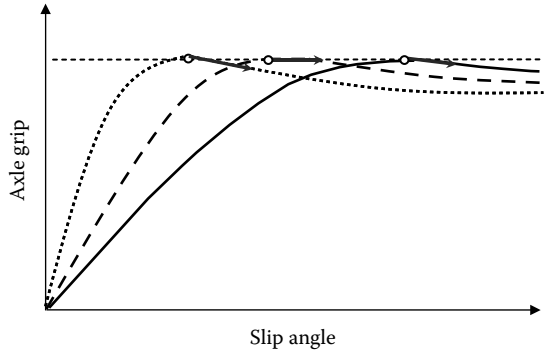


FIGURE 37.23 Different types of limit grip.

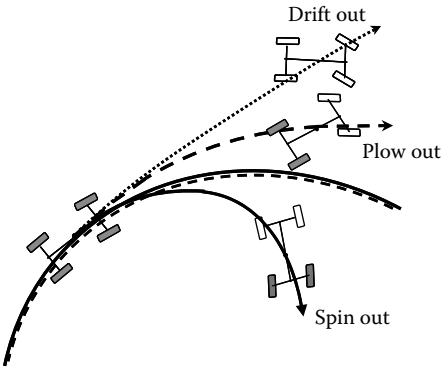


FIGURE 37.24 Limit-grip behavior.

Figure 37.24 presents three different cases of limit-grip behavior depending on the limits of cornering grip such as (a) *plow out*, (b) *spin out*, and (c) *drift out*. The plow out or spin out is caused due to the yaw moment of unbalanced cornering forces at front and rear axles. Specifically speaking, the plow out or spin out does wholly depend on the less grip of the front and rear axles relative to each other, respectively. In case the cornering forces are ideally well balanced, a vehicle does not get either plow out or spin out. It eventually gets drift out at a little higher grip limit that depends on the overall cornering grip of the vehicle.

37.2.3.4 Controllability at Limit

Controllability at limit is an overall impression on how easy to control a vehicle at grip limit. It is ideally preferable to have flat limit grip without its sudden drop so that in advance a driver can prepare for the next situation (refer to Figure 37.23). As slip angle increases, in practice, eventually the limit grip decreases. Some tires behave little oversteer at the grip limit, and it may be easy for a specially trained driver to control a vehicle with them.

37.2.3.5 Driving Power

Driving power is simply a traction ability to speed up while powering on in the exit zone of a corner. With good driving power, a vehicle can rhythmically follow the ideal line of the path in a handling circuit. It provides the perception of *agility* as well as a pleasure of sporty feeling.

37.2.4 Stability

During severe handling maneuver beyond grip limit, a vehicle may get into an unstable condition. With a driver's proper control, in this situation, some vehicles can recover (or converge) to a stable driving condition. Some others diverge to various accidents without recovery. Stability is an ability to make the vehicle come back to a next stable turn or lane change. It is represented by its six characteristics such as (a) *yaw stability*, (b) *rear axle stability*, (c) *roll stability*, (d) *braking stability*, (e) *rollover stability*, and (f) *bump steer.*

37.2.4.1 Yaw Stability

Yaw stability is an ability to keep down the yaw oscillation of a vehicle. The yaw oscillation is principally generated due to an unbalance of the lateral compliances between front and rear axles. In case the roll compliances of front and rear axles are unbalanced with their lateral compliances, the yaw oscillation may be also produced. For instance, if vehicle body roll motion has some phase lag behind vehicle path change, the yaw instability also occurred.

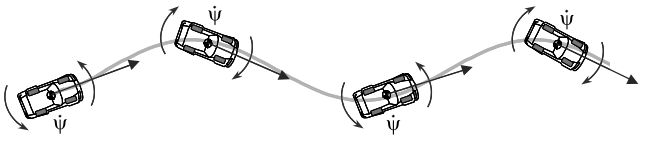


FIGURE 37.25 Yaw stability.

Depending on yaw damping, the cycle number of the yaw oscillation is determined. The fewer the cycles, the better. In practice, the yaw stability is assessed primarily using a perceptual measurement of the oscillation of vehicle path as shown in Figure 37.25.

37.2.4.2 Rear Axle Stability

Rear axle stability is an ability to restrict the sideslip oscillation of a vehicle. This oscillation has occurred due to the lack of rear axle grip relative to front axle grip. For good rear axle stability, the magnitude of the sideslip oscillation should be small enough. In addition, it should be easy to recover. As shown in Figure 37.26, the rear axle stability is evaluated mainly through perceptually measuring the oscillation of vehicle attitude.

37.2.4.3 Roll Stability

Roll stability is an ability to depress the roll oscillation of a vehicle. The excessive roll oscillation is mainly due to an improper balance of the roll compliance of the vehicle. Both the magnitudes of roll oscillation and roll damping are important for roll stability (refer to Figure 37.27). In practice, a driver's subjective rating on the roll stability is additionally influenced by vehicle pitch oscillation. The driver cannot completely separate the vehicle roll from the vehicle pitch only using one's perception without an objective measurement. The roll stability is assessed by using a perceptual measurement on the oscillation of vehicle body motion.

37.2.4.4 Braking Stability

Braking stability is an ability to stop a vehicle without any disturbance of vehicle attitude or path while full braking in a turn. With bad braking stability, a vehicle stops with the misalignment of vehicle heading direction along an intended path. The braking stability is assessed by perceptually evaluating the disturbance(s) of the vehicle attitude and/or path.

37.2.4.5 Rollover Stability

Rollover stability is an ability to restrain a vehicle from its rollover motion (refer to Figure 37.28). In general, the rollover gets into a fatal accident. Therefore, these days, it is a hot issue to automotive

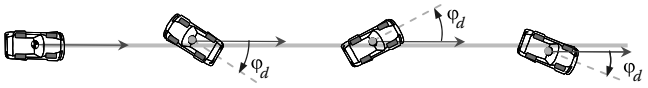


FIGURE 37.26 Rear axle stability.

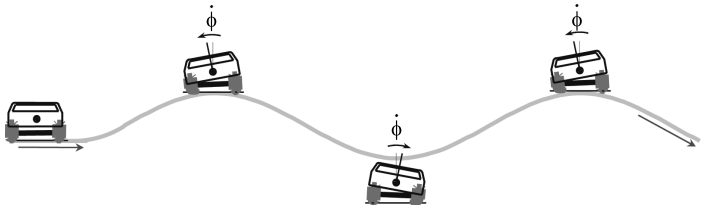


FIGURE 37.27 Roll stability.

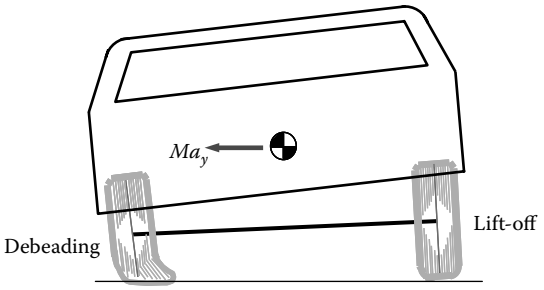


FIGURE 37.28 Rollover stability.

makers as well as tire makers. In order to assess the rollover stability, both the debead of a tire(s) from a rim(s) and the separation of a tire(s) from road surface are subjectively or objectively checked through a series of specially designed rollover tests.

37.2.4.6 Bump Steer

While a vehicle is running over bumps, its toe is constantly changed due to the jounce and rebound of each front suspension (refer to Figure 37.29). In this case, steer instability may occur. In that case, the vehicle cannot follow an indented path in a turn. Instead, the vehicle follows disturbed paths, and it requires continuous steering corrections to come back to the intended path. This steer instability is called bump steer.

37.2.5 Comfortability

Comfortability belonged to the areas of both vehicle dynamics and noise, vibrations, and harshness (NVH). The vehicle dynamics deals only the vibrations of sprung and unsprung masses. On the other hand, the NVH represents the vibrations of flexible components. In this handbook, the comfortability representing ride performance is limited only within the scope of the vehicle dynamics. The comfortability is composed of (a) *primary ride*, (b) *secondary ride*, and (c) *impact ride*.

37.2.5.1 Primary Ride

Primary ride is the vibratory characteristic of a sprung mass associated with the principal stiffnesses of tires occurred while driving on a wavy road. Its frequency ranges up to about 5 Hz, which includes some frequency band inducing motion sickness. The specific range is wholly dependent on the size of a vehicle. In general, the primary ride is principally represented by *body bounce*. But it may be specifically categorized into several zones of frequency ranges.

37.2.5.2 Secondary Ride

Similar to the primary ride, secondary ride is the vibratory characteristic of unsprung masses supported with the principal stiffnesses of tires generated while driving a rough road. Its frequency ranges from about 5 to 25 Hz. Its range is also dependent on the size of a vehicle. Similar to the

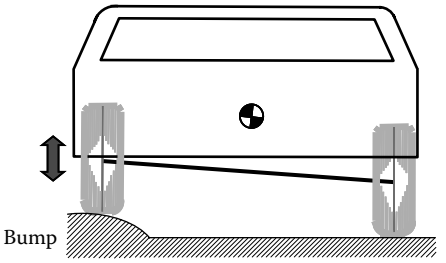


FIGURE 37.29 Bump stability.

primary ride, the secondary ride is generally represented by *wheel hop*. However, it can be specifically divided into several zones of frequency ranges.

37.2.5.3 Impact Ride

Impact ride is wholly determined by an impact shock(s) due to the enveloping characteristic(s) of a tire(s) while driving over various impact objects. In general, it is called *impact harshness*. Its frequency usually ranges from about 25 to 60 Hz (at most 100 Hz). The impact ride does not include the local vibratory characteristic of tire components, which produce higher-frequency vibrations belonging to an NVH scope. In reality, a driver's subjective assessment on the impact ride is also influenced by audible high frequencies in addition to tactical low frequencies.

The impact ride is generally characterized as (a) *impact magnitude*, (b) *impact sharpness*, and (c) *after-impact damping*. The impact magnitude is just the size of an impact shock. The impact sharpness is the level of sharpness (or acuteness) of the impact shock, which ranges from round to sharp. The after-impact damping is the amount of damping to kill the vibration of a tire itself after impacting. The smaller the magnitude, the less the sharpness, or the more the damping, the better.

37.3 SUBJECTIVE TEST

Dry handling and ride performances are assessed through five *maneuvers* such as (a) *straight ahead*, (b) *steering*, (c) *lane changing*, (d) *cornering*, and (e) *riding* as shown in Figure 37.30 [40–43]. Each maneuver is composed of several *tests*. For the purpose of evaluating wet handling, however, only some tests are usually selected dependent on each company. In Table 37.1, a cross-functional relationship between subjective assessment and subjective test is illustrated using a cross-functional matrix of capabilities and maneuvers. In Tables 37.2 through 37.6, these matrices are shown in details as cross functions of characteristics against tests.

For each test, both *curb weight with a driver* (or *unladen*) and *gross vehicle weight (GVW)* (or *fully laden*) conditions are in general used. However, it is wholly depending on each automaker or tire company. Each test condition and procedure is individually described as follows.

37.3.1 Straight-Ahead Maneuver

A straight-ahead maneuver is composed of four tests of straight-ahead driving with (a) *free steering*, (b) *fixed steering*, (c) *powering on/off*, and (d) *braking*. These tests are designed for the purpose of assessing straightability.

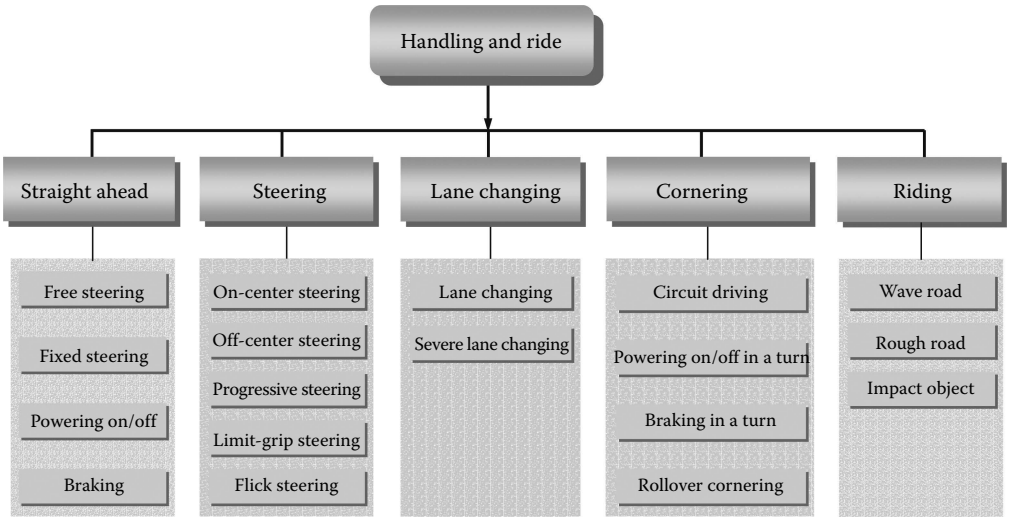


FIGURE 37.30 Subjective test for the assessment of dry handling and ride performances.

TABLE 37.1
Cross-Functional Matrix between Capacities and Maneuvers

Capability	Maneuver	Straight Ahead	Steering	Lane Changing	Cornering	Riding
	Straightability	✓				
	Steerability		✓		✓	
	Controllability			✓	✓	
	Stability		✓	✓	✓	
	Comfortability					✓

TABLE 37.2
Cross-Functional Matrix between Characteristics and Straight-Ahead Tests

		Straight Ahead			
Performance	**Maneuver**	**Free Steering**	**Fixed Steering**	**Powering On/Off**	**Braking**
Straightability	Residual pull	✓			
	Running straight		✓		
	Torque steer			✓	
	Braking straight				✓

TABLE 37.3
Cross-Functional Matrix between Characteristics and Steering Tests

		Steering				
Performance	**Maneuver**	**On-Center Steering**	**Off-Center Steering**	**Progressive Steering**	**Limit-Grip Steering**	**Flick Steering**
Steerability	On-center	✓				
	Steering response		✓			
	Linearity			✓		
	Steering precision					
	Steering angle					
	Steering addition				✓	
	Steering effort		✓			
	Returnability					✓
Stability	Yaw stability					✓
	Rear axle stability					✓
	Roll stability					✓
	Braking stability					
	Rollover stability					
	Bump steer					

TABLE 37.4
Cross-Functional Matrix between Characteristics and Lane-Changing Tests

		Lane Changing	
Performance	**Maneuver**	**Lane Changing**	**Severe Lane Changing**
Controllability	Balance FA/RA	✓	✓
	Power-off reaction		
	Grip limit	✓	✓
	Controllability at limit	✓	✓
	Driving power		
Stability	Yaw stability	✓	✓
	Rear axle stability	✓	✓
	Roll stability	✓	✓
	Braking stability		
	Rollover stability		
	Bump steer		

TABLE 37.5
Cross-Functional Matrix between Characteristics and Cornering Tests

		Cornering			
Performance	**Maneuver**	**Circuit Driving**	**Powering On/ Off in a Turn**	**Braking in a Turn**	**Rollover Cornering**
Steerability	On-center				
	Steering response				
	Linearity				
	Steering precision	✓			
	Steering angle	✓			
	Steering addition				
	Steering effort	✓			
	Returnability	✓			
Controllability	Balance FA/RA	✓	✓		
	Power-off reaction		✓		
	Grip limit	✓			
	Controllability at limit	✓	✓		
	Driving power	✓	✓		
Stability	Yaw stability				
	Rear axle stability		✓		
	Roll stability				
	Braking stability			✓	
	Rollover stability				✓
	Bump steer	✓			

TABLE 37.6
Cross-Functional Matrix between Characteristics and Riding Tests

		Riding		
Performance	**Maneuver**	**Wave Road**	**Rough Road**	**Impact Object**
Comfortability	Primary ride	✓		
	Secondary ride		✓	
	Impact ride			✓

37.3.1.1 Free Steering in Straight Ahead

A free-steering-in-straight-ahead test is targeted to check whether or not a vehicle can run in straight ahead with free steering at constant speed. Its purpose is to examine the *residual pull* characteristic of the suspension system, steering system, and tires of a vehicle (refer to Figure 37.8). The summary of test condition and procedure is given as

Driving speed: Constant speed usually at 100 kph with a high gear.
Steering input: Free steering at the center steering.
Procedure: Start free steering from the exact center steering point after removing a static friction effect.

37.3.1.2 Fixed Steering in Straight Ahead

A fixed-steering-in-straight-ahead test is designed to find the *running straight* characteristic of a vehicle to run in straight ahead without any drift, wander, or sloppiness at constant speed (refer to Figure 37.9). This lateral deviation is caused due to suspension sensitiveness against road irregularity or wind change. For this test, fixed steering is employed as

Driving speed: Constant speed up to limit.
Steering input: Fixed steering at the center steering.
Procedure: Start fixed steering from the exact center steering point after removing a static friction effect.

37.3.1.3 Powering On/Off in Straight Ahead

The purpose of a powering-on/off-in-straight-ahead test is to evaluate the *torque steer* characteristic of a vehicle. The torque steer is subjectively determined by the degree of drifting motion of the vehicle with fixed steering (refer to Figure 37.10). In general, the vehicle drifts to an adjacent lane with full power on and returns to the original lane with instantaneous power off.

In order to find the defect of a power train system, second transmission gear is employed. On the other hand, a higher gear is used to examine the flaw of tires:

Driving speed: 40 kph with second gear and 80 kph with a higher gear.
Steering input: Fixed steering at the center steering.
Procedure: Apply full power on and then suddenly lift throttle off in straight ahead.

37.3.1.4 Braking in Straight Ahead

Through a braking-in-straight-ahead test with hands-on steering, the *braking straight* characteristic of a vehicle is subjectively measured after complete stop. The braking straight is usually represented

by the degree of deviation of vehicle attitude. Soft hands-on steering should be applied so that steering disturbance is also detected at the end of braking process:

Driving speed: 100 kph or normal driving speed.
Steering input: Soft hands-on steering at the center steering.
Procedure: Apply full braking in straight ahead until complete stop with VDC (also called ESP or ESC) turned off.

37.3.2 Steering Maneuver

In order to evaluate the steering behavior of a vehicle, steering maneuver is specially designed. In general, it consists of five steering tests such as (a) *on-center steering*, (b) *off-center steering*, (c) *progressive steering*, (d) *limit-grip steering*, and (e) *flick steering*. For these tests, basically both clockwise and counterclockwise directions of steering input are applied. These processes are repeated a few times until subjective assessment is complete.

In this handbook, "*perceptual*" is used to describe a ramp, step, impulse, or sinusoidal steering input, so that a perceptual one is distinguished from a mechanical one. For example, a perceptual ramp steering input is not a real ramp steering input, but only it is perceived like a ramp steering input.

37.3.2.1 On-Center Steering

Through an on-center steering test, *on-center feel* at and around the center of steering is subjectively assessed (refer to Figures 37.12 and 37.13). Since the on-center feel is a kind of microscopic property, it is very important that vehicle speed should be kept constant during the whole test process. In addition, the application of an on-center steering input must be very gentle and precise in order to feel the on-center steering behavior of a vehicle.

For this test, a slow perceptual ramp steering input with its miniature magnitude from the steering center to one direction is applied while driving in straight ahead. After releasing a steering handwheel in order to return to straight running again, the same amount of the steering input from the steering center to the other direction is employed. In case these steps are repeated for several cycles, the steering input looks like a weave steering input, which eventually produces the profile of steering torque:

Driving speed: Constant speed up to limit.
Steering input: Miniature *slow perceptual ramp* steering.
Procedure: Apply a steering input from the steering center in straight ahead.

37.3.2.2 Off-Center Steering

The purpose of an off-center steering test is to assess (a) *steering response* and (b) *steering effort* while driving both in straight ahead and in a turn (refer to Figures 37.14 and 37.17). This test contains two different test methods. The first method requires a small perceptual step steering input while driving in straight ahead. In addition, the second method employs the steering input in a turn. It is important that this test requires very clear steering stop just after the application of the perceptual step steering input in order to measure the time delay of vehicle turning reaction after the steering input:

Driving speed: Constant speed up to limit.
Steering input: Small *perceptual step* steering.
Procedure: Apply a steering input both in straight ahead and in a turn.

37.3.2.3 Progressive Steering

While slowly crossing double or triple lanes (the wider, the better), a progressive steering test is carried out. Its purpose is to evaluate the *linearity* characteristic of vehicle path reaction against a

gradually incremental steering input (refer to Figure 37.15). This test requires a big or middle perceptual ramp steering input while driving in straight ahead:

Driving speed: Constant speed up to limit.
Steering input: Big or middle *perceptual ramp* steering.
Procedure: Apply a steering input in straight ahead.

37.3.2.4 Limit-Grip Steering

A limit-grip steering test is designed to evaluate *steering addition* to find whether or not tires can produce an additional cornering force during limit-grip cornering in a handling circuit. Sometimes, a driver wants to have some more turn(s) in the middle of a corner in order to drive an understeer behaved vehicle back to an intended path or overcome any uncertainty of driving condition. For this test, a small perceptual ramp steering input is additionally applied at the limit grip in a corner. Thereafter, vehicle path change is observed:

Driving speed: Constant speed at limit.
Steering input: Small *perceptual ramp* steering.
Procedure: Apply a steering input at grip limit in a corner.

37.3.2.5 Flick Steering

When a vehicle is running in straight ahead, a flick steering input is applied with 90° of steering handwheel angle away from the steering center. Then, immediately, the vehicle gets imposed into an unstable situation with yaw oscillation. In case the yaw oscillation is damped out, the vehicle can get back to a stable driving condition. Otherwise, the yaw oscillation may be amplified to eventually spin out as shown in Figure 37.24. This recovery characteristic is dependent on the self-aligning moment, front–rear lateral compliance balance, roll compliance, and rear axle grip of tires including corresponding vehicle subsystems.

Through the flick steering test, the recovery characteristic is subjectively assessed in terms of (a) *returnability*, (b) *yaw stability*, and (c) *roll stability* (refer to Figures 37.18, 37.25, and 37.27). In the case of critically bad rear axle grip, (d) *rear axle stability* may be supplementarily evaluated (refer to Figure 37.26). Especially, the test is designed to mainly evaluate the yaw stability distinctively against the rear axle stability. For this test, a perceptual impulse (or flick) steering input is applied such that 90° of steering handwheel angle is turned from the steering center as quickly as possible while driving in straight ahead and then immediately the steering handwheel is released free. Afterward, observe the returning motion of the steering handwheel and vehicle reaction as well:

Driving speed: Constant speed up to limit.
Steering input: 90° of *perceptual impulse* steering.
Procedure: Apply a steering input from the steering center and then immediately release the steering handwheel in straight ahead.

37.3.3 Lane-Changing Maneuver

In practice, lane changing is most frequently operated as a handling maneuver in highways and public roads. Depending on its severity, the lane changing induces the modes of controllability and stability. This maneuver is usually composed of two tests of (a) *lane changing* and (b) *severe lane changing.*

37.3.3.1 Lane Changing

The purpose of a lane-changing test is in general to pass over a vehicle appeared in front at a highway as well as a freeway as shown in Figure 37.31. While lane changing, both controllability and stability are mainly evaluated by checking transient understeer and oversteer tendencies and stability characteristics. Through this test, (a) *balance FA/RA*, (b) *grip limit*, (c) *controllability at limit*,

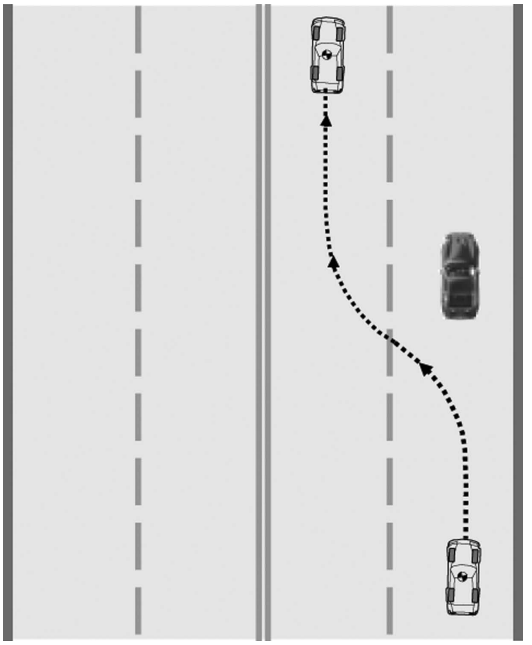


FIGURE 37.31 Lane changing.

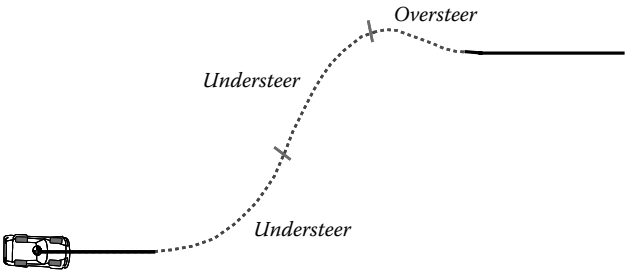


FIGURE 37.32 Steer characteristic during lane changing.

(d) *yaw stability*, (e) *rear axle stability*, and (f) *roll stability* are assessed (refer to Figures 37.19, 37.20, 37.23, 37.24 through 37.27).

The lane-changing test basically contains two consecutive turns in opposite directions in order to cross a single lane. For example, as shown in Figure 37.32, a vehicle is initially turning left and then is turning right in order to run straight in the next left lane. In general, a vehicle behaves *transient understeer* in the first and second turns. *Transient oversteer* is subjectively assessed at the end of the lane changing as shown in Figure 37.32. In order to achieve two consecutive turns, three phases of steering handwheel operations are needed such as (a) left-hand-side steering, (b) steering from the left-hand side to the right-hand side, and (c) left-hand-side steering to the center of steering.

For this test, the angle of a steering handwheel ranges between around ±45°. In addition to this constrained steering input, there exists another constraint that a test engineer should keep vehicle lateral movement within the next lane. Depending on the reaction of vehicle path change in response to each steering application, the speed of the next steering application should be properly adjusted. In case the vehicle has bad lane-change behavior, it requires an additional steering input to come back to the path:

Driving speed: Constant speed up to limit.
Steering input: 45° of *perceptual sinusoidal* steering.

Procedure: Apply three consecutive steering input with constant throttle ideally as

$$\text{center} \xrightarrow{\pm 45^\circ} \frac{R45^\circ}{L45^\circ} \xrightarrow{\mp 90^\circ} \frac{L45^\circ}{R45^\circ} \xrightarrow{\pm 45^\circ} \text{center}$$

37.3.3.2 Severe Lane Changing

For the purpose of avoiding a vehicle not to run over an object, such as a child or a ball, suddenly appeared in front, severe lane changing is required as shown in Figure 37.33. In this kind of an emergency situation, a driver normally tends to instantly lift throttle off. When only a single lane exists in each direction, once the vehicle moves to the next lane, then it should immediately return to the original lane. Since the vehicle totally crosses two lanes, more often, severe lane changing is called double-lane changing.

Through the severe lane-changing test, (a) *balance FA/RA*, (b) *grip limit*, (c) *controllability at limit*, (d) *rear axle stability*, and (e) *roll stability* are mainly assessed (refer to Figures 37.19, 37.20, 37.23, 37.24, 37.26, and 37.27). In case a front–rear lateral or roll compliance balance is very bad, (f) *yaw stability* is additionally evaluated (refer to Figure 37.25). Particularly, the rear axle stability is distinctively evaluated through this test, while the yaw stability is assessed through a flick steering test.

The severe lane changing test is basically composed of three consecutive turns. For example, as shown in Figure 37.34, it is assumed that a vehicle turns a left lane and then backs to the original lane. The severe lane-change process is done through (a) the first turn to about halfway to the left lane, (b) the second turn to the left lane and then continuously back to the original lane, and (c) the last turn to run straight in the original lane. The vehicle usually behaves *transient understeer* in the beginning of every turn, but it may behave *transient oversteer* in the first turn due to sudden power off even at 80 kph. In general, the *transient oversteer* is subjectively evaluated at the ends of the second and third turns.

In order to achieve these three turns, five phases of steering handwheel operations are needed such as (a) initially left-hand-side steering, (b) steering from the left-hand side to the right-hand side, (c) fixed steering with about 90° of steering handwheel angle for a steady-state turn, (d) left-hand-side steering to make vehicle path straight, and then (e) right-hand-side steering back to the center of steering. In addition, some more steering corrections may be needed depending on vehicle performance. The steering handwheel angle is supposed to range between about ±90°.

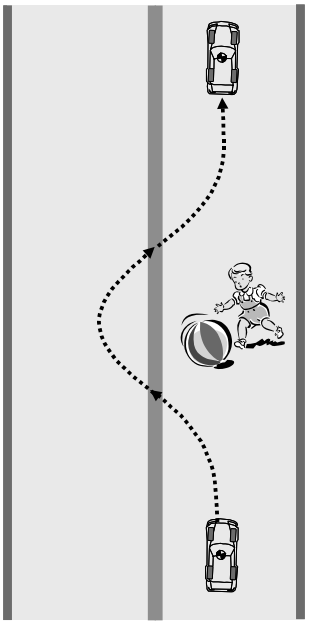


FIGURE 37.33 Severe lane changing.

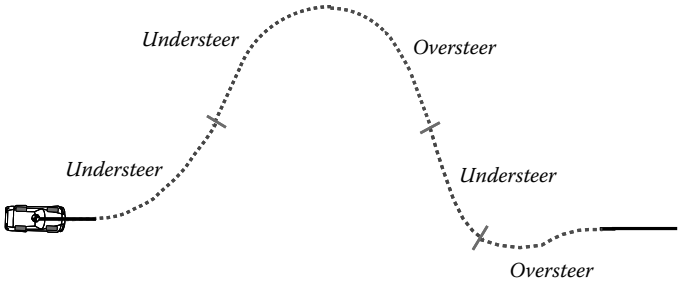


FIGURE 37.34 Steer characteristic during severe lane changing.

Similar to the lane changing, there exist two constraints of the steering handwheel angle and vehicle lateral movement restricted only within two adjacent lanes. The speed of each steering application and the duration of the fixed steering are appropriately determined based on vehicle path reaction in response to the previous steering application:

Driving speed: 80 kph.
Steering input: 90° of *perceptual sinusoidal* steering.
Procedure: Apply four consecutive steering inputs starting with instant power off ideally as

$$\text{center} \xrightarrow{\pm 90^\circ} \begin{matrix} R90^\circ \\ L90^\circ \end{matrix} \xrightarrow{\mp 180^\circ} \begin{matrix} L90^\circ \\ R90^\circ \end{matrix} \xrightarrow{\pm(90^\circ + \Delta\delta_{sw})} \begin{matrix} R\Delta\delta_{sw} \\ L\Delta\delta_{sw} \end{matrix} \xrightarrow{\pm \Delta\delta_{sw}} \text{center}$$

37.3.4 Cornering Maneuver

Cornering maneuver is just normally driving in corners such as an entrance to a highway, an exit from a highway, or an interchange between highways or public roads. Through this maneuver, (1) steerability, (2) controllability, and (3) stability are usually investigated. The maneuver consists of four tests of (a) *circuit driving*, (b) *powering on/off in a turn*, (c) *braking in a turn*, and (d) *rollover cornering*. During cornering maneuver, sometimes a vehicle plows out, spins out, or drifts out of a course. In this situation, it is required to control the vehicle to come back to the course in order to avoid any accident.

In order to control a vehicle while sliding at or beyond grip limit, at least the effective heading and moving directions of the vehicle itself, front tires, and rear tires must be perceived at any instant. In a transient turn, these two directions of each component are normally seldom coincided. The moving directions of the vehicle, front tires, and rear tires are dependent on each other. The effective heading direction of the rear tires is about along that of the vehicle in the case of a front-wheel-steering car. However, the effective heading direction of the front tires is completely controlled only by a steering input.

For the purpose of explaining the mechanism of how to control a vehicle while sliding, two cases are discussed here:

1. Sliding at or beyond grip limit
   When a vehicle plows out, spins out, or drifts out at or beyond grip limit as shown in Figure 37.35, steer a steering handwheel to align the effective heading direction of front tires along the moving direction of a vehicle so that slip angles at the front tires become about zero. For that purpose, continuously adjust a steering handwheel to make the self-aligning moment of a front axle be nearly zero. Once the vehicle is controlled from sliding, steer the steering handwheel appropriately to follow an intended path.

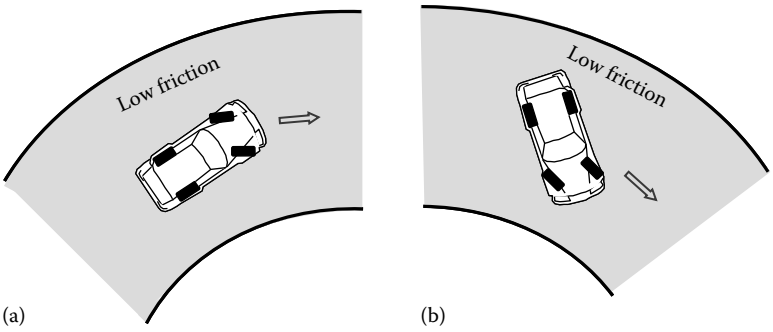


FIGURE 37.35 Control of a vehicle by aligning the heading direction of front tires along vehicle moving direction. (a) No sliding and (b) spin out.

2. Crossing from low friction to high friction
   Within low friction surface, follow exactly the same procedure of the previous case as shown in Figure 37.35. When front tires are crossing from low-friction surface to high-friction surface such as from wet surface to dry surface as shown in Figure 37.36a, hold a steering handwheel to keep the front tire heading direction along the vehicle moving direction. As soon as the front tires have entered the high-friction surface as shown in Figure 37.36b, immediately

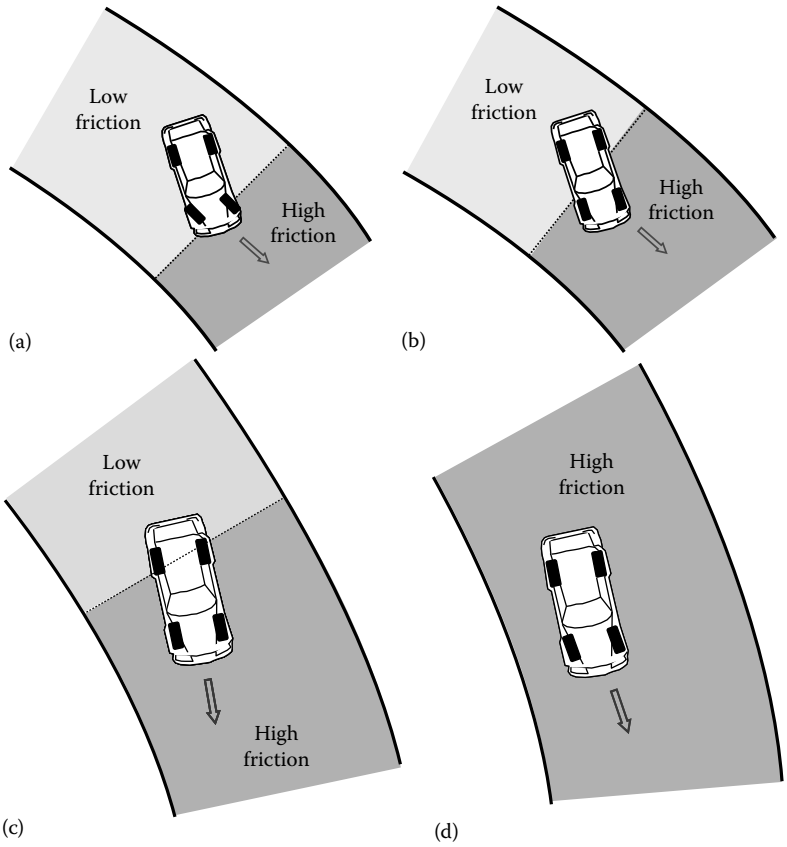


FIGURE 37.36 Control of a vehicle by aligning front tires while crossing from low-friction surface to high-friction surface. (a) Crossing of front tires, (b) just after crossing of front tires, (c) crossing of rear tires, and (d) following an intended path.

readjust the steering handwheel to align the front tire heading direction along the vehicle heading direction in order to make a turn.

As a result, the vehicle moving direction becomes aligned along about the vehicle heading direction as much as possible. In other words, the vehicle moving direction becomes along close to the effective heading direction of rear tires so that slip angles at the rear tires become as small as possible. When the rear tires are crossing from the low-friction surface to the high-friction surface as shown in Figure 37.36c, just keep the same steering handwheel angle so that slip angles at the rear tires become nearly zero. Immediately after the rear tires have entered the high-friction surface as shown in Figure 37.36d, follow an intended path using normal steering maneuver.

It is noted that at the low-friction surface, a longitudinal load shift has not much occurred due to power on/off or braking in a turn so that the load shift does not much induce the loss of cornering grip (refer to Figures 37.3b and 37.4b). Instead, the effect of a friction ellipse concept is much significant on the cornering grip loss (refer to Figure 37.5). At the high friction, however, the effect of the load transfer on the cornering grip loss is much bigger than that of the friction ellipse concept.

37.3.4.1 Circuit Driving

A circuit driving test is mainly targeted to make an overall assessment on path following characteristic that is also called *line trace*. This test in a handling circuit is also called a *road course test*. Through the test, (a) *steering precision*, (b) *steering angle*, (c) *steering effort*, (d) *returnability*, (e) *balance FA/RA*, (f) *grip limit*, (g) *controllability at limit*, and (h) *driving power* are subjectively evaluated (refer to Figures 37.16 through 37.20, 37.23, and 37.24). In order to do the circuit driving test successfully, the maneuvers of braking, gear changing, and steering must be properly done.

Figure 37.37 presents a standard cornering procedure in a handling circuit. An ideal path in a corner is always pursued based on the principle of "*creating the biggest turning radius*," so that the maximum exiting vehicle speed is achieved. For that purpose, a driver must directly target to drive ahead to three points: (1) an entering point in straight ahead or from the exiting point of the just previous corner, (2) an apex (or a clipping point) directly headed from the entering point, and (3) an exiting point directly forwarded from the apex. The cornering procedure contains four phases such as (a) approaching a corner, (b) turning in a corner, (c) holding in a steady-state turn, and (d) exiting in a corner.

While approaching a corner, entering vehicle speed should be appropriately adjusted through braking. During turning in a corner, a steering handwheel angle is progressively increased with constant throttle at limit speed. Afterward, steady-state cornering is carried out until reaching the apex. In order to exit in a corner from the apex to the exiting point, full power on is generally applied to get the maximum vehicle speed. During this exiting phase, the steering handwheel

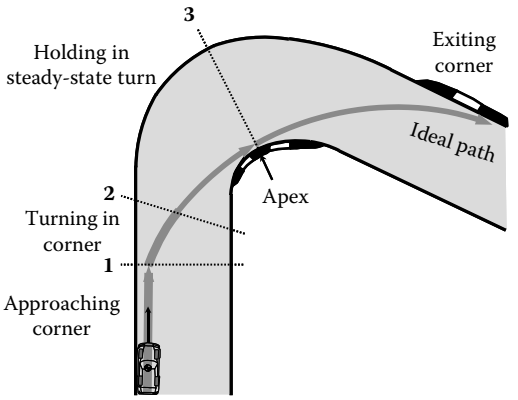


FIGURE 37.37 Circuit driving.

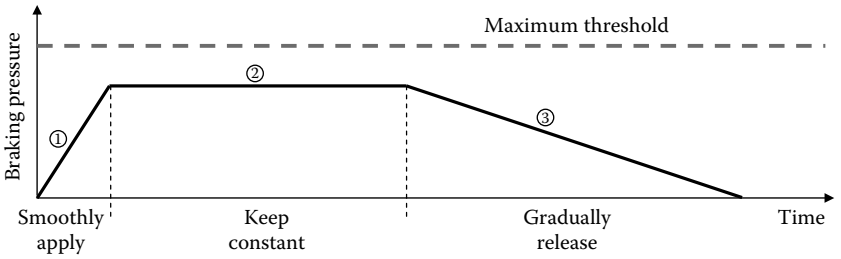


FIGURE 37.38 Braking process.

angle is gradually released by using self-aligning moment, which makes a steering handwheel naturally return to a steering center point.

Before entering a corner, a proper braking process is first carried out to reduce vehicle speed. As shown in Figure 37.38, the process is composed of three steps: (1) smoothly apply braking pressure, (2) keep constant braking pressure until getting intermediate desired speed, and (3) gradually release the braking pressure to get the final desired speed. The same level of constant braking pressure is targeted before entering all corners in a handling circuit. The application of braking just before running over any impact object should be avoided.

While braking, a gear downshift is complete. When the gear downshift is carried out from a higher gear to a lower gear, it is necessary to raise engine rpm so that there causes no shock to a drive line. A professional way to do the gear downshift is called *heel and toe* as shown in Figure 37.39. For that purpose, the right foot of a driver is on both a brake pedal and a throttle pedal, while the left foot is on a clutch pedal. Specifically speaking, while the toe of the right foot is still on the brake pedal, its heel is turned to the throttle pedal. Then, rotate the heel to flip the throttle pedal to increase the engine rpm while keeping braking with the toe. Afterward, the gear downshift can be smoothly done without any shock.

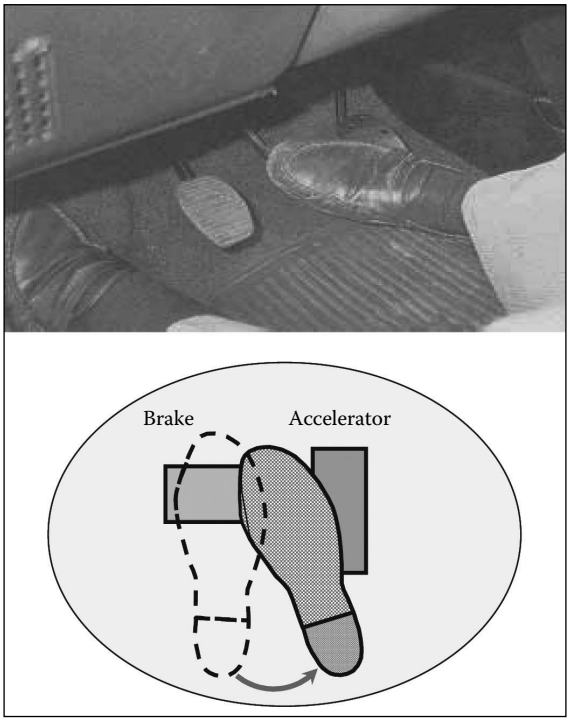


FIGURE 37.39 Heel and toe.

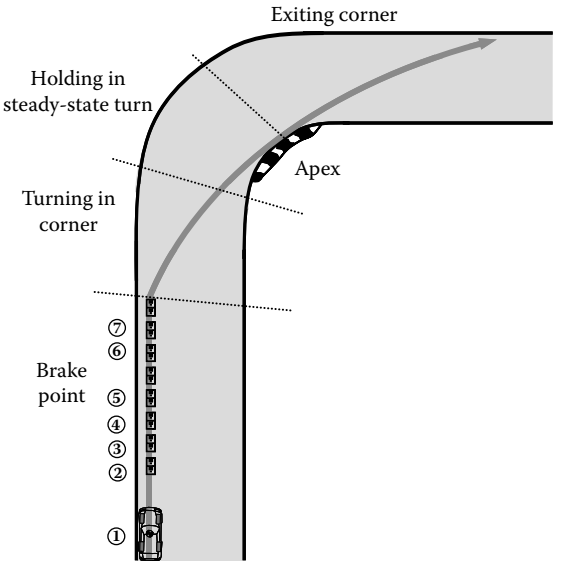


FIGURE 37.40 Gear downshift and other procedure during circuit driving.

As shown in Figure 37.40, the process of gear downshift before cornering is summarized as

1. Approach to a corner
2. Start to brake
3. Declutch and simultaneously gear shift down while keeping braking
4. Flip throttle on to match engine rpm while keeping braking, or keep throttle on until the rpm is adjusted
5. Immediately release clutch
6. Keep constant throttle on up to the entry point
7. Turn in while keeping constant throttle on

In case the engine rpm matching is not appropriate in Step 4, a vehicle shock is induced in Step 5. If a clutch is released too early, a forward shock is caused due to too high engine rpm. If it is delayed to release the clutch, on the other hand, a rearward shock occurred due to not enough engine rpm.

For the purpose of series of steering applications to follow an intended path in a handling circuit, the proper operation of a steering handwheel is necessary. The angle between two hands on a steering handwheel is desired to keep about 150° in any situation as shown in Figures 37.41 and 37.42. Each arm is supposed to keep bent about 120° of angle between its upper and lower parts. It is important to note that at every corner, a hands-on steering application is required in order to be gradually increased and be smoothly released during turning in and exiting in a corner, respectively.

In case of steering to the right-hand side shown in Figure 37.41, the left and right hands are turned from (a) and (b) of Phase 1 to (c) and (d) of Phase 2, respectively. Afterward, the left hand is relocated to (e) of Phase 3, and then the right hand is shifted to (f) of Phase 4. For the left-hand-side steering, the similar process to the right-hand-side steering is applicable as shown in Figure 37.42. The right hand is located from (b) of Phase 1 to (d) of Phase 2 and then to (e) of Phase 3. The left hand is moved from (a) of Phase 1 to (c) of Phase 2 and finally to (f) of Phase 4.

The condition and process of a circuit driving test are summarized as

Driving speed: Limit speed.
Steering input: Path following steering.

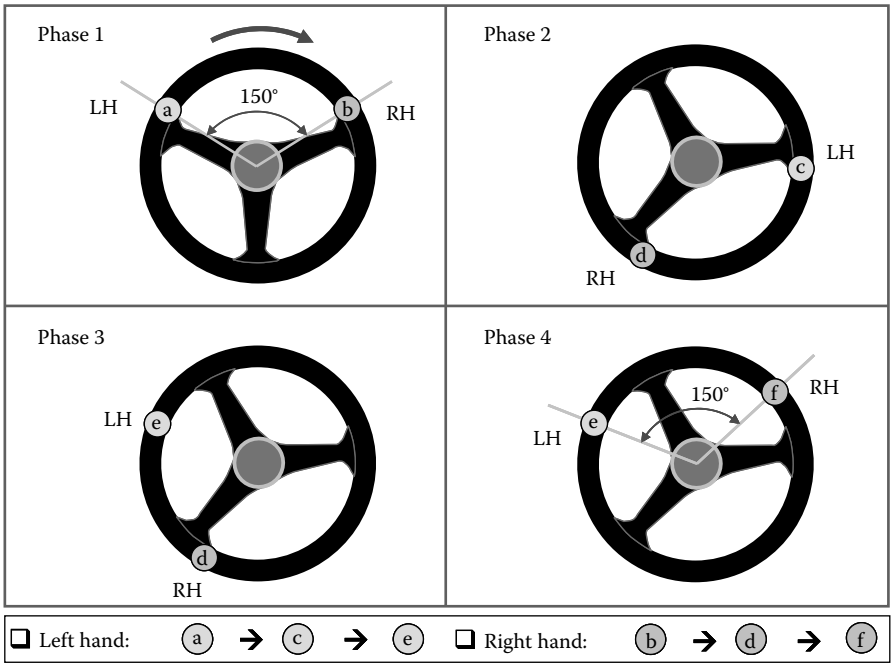


FIGURE 37.41 Steering to right-hand side.

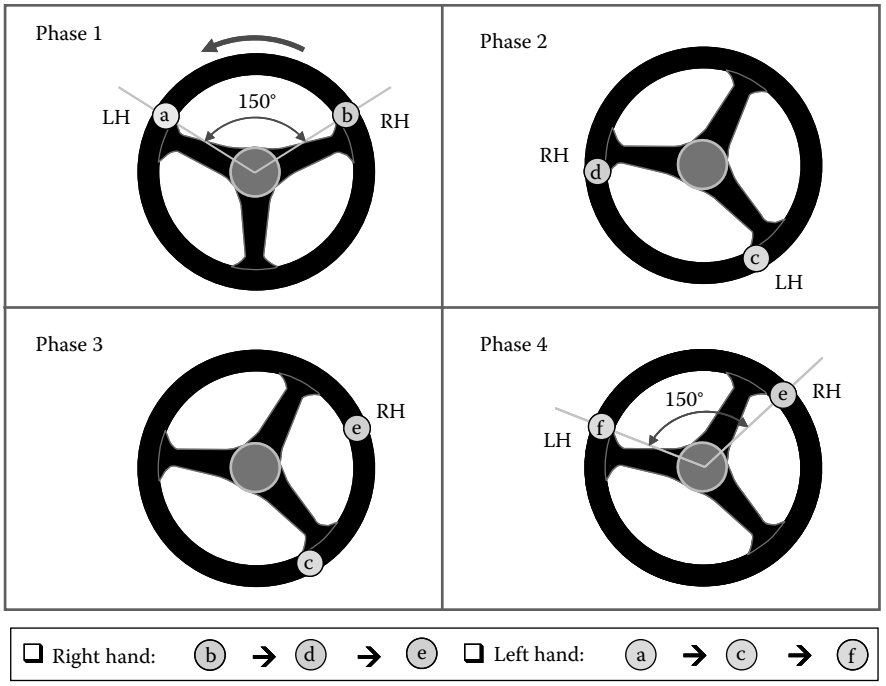


FIGURE 37.42 Steering to left-hand side.

Procedure: Make a corner turn as follows (refer to Figure 37.40):

1. Brake and downshift to a lower gear in straight ahead before entering a corner
2. Approach in a corner toward an outer curve line
3. Follow along the outer curve line with constant throttle
4. Turn toward an apex (or a clipping point) with constant throttle
5. Hold a steady-state turn with fixed steering and constant throttle
6. Exit in the corner toward an outer curve line with power on

37.3.4.2 Powering On/Off in a Turn

While we are entering in a corner from straight driving at high speed, from time to time, we realize that our entering vehicle speed is too high to make a safe turn. In that situation, in general, a normal driver immediately lifts throttle off in order to reduce the speed. Then, there exists an instantaneous load transfer from a rear axle to a front axle. This load transfer causes the loss of rear axle grip and instead adds more to front axle grip, which sometimes makes serious oversteer tendency due to a rear axle's swing out (refer to Figure 37.22).

The main purpose of a powering-on/off-in-a-turn test is to evaluate *controllability* to overcome a *load transfer effect* from a rear axle to a front axle when momentarily throttle changing in a turn. In addition, *stability* is also assessed in the worst case. For that purpose, (a) *balance FA/RA*, (b) *power-off reaction*, (c) *controllability at limit*, (d) *driving power*, and (e) *rear axle stability* are subjectively measured (refer to Figures 37.19, 37.20, 37.22, 37.24, and 37.26). In the case of a front-wheel-driving car, in general, full power on increases understeer tendency, but it reduces understeer tendency with a rear-wheel-driving car.

A sudden transition from full power on to power off causes an impulsive load shift, which results in an instantaneous decrease and an increase to the cornering forces of a rear axle and a front axle, respectively. The change of the cornering forces does quickly increase a yaw moment momentarily producing big oversteer. It should be considered that steering corrections are needed to bring the vehicle back on a course:

Driving speed: Limit speed.
Steering input: Fixed steering.
Procedure: Apply full power on before limit speed in a turn and momentarily power off at the limit speed.

37.3.4.3 Braking in a Turn

A braking-in-a-turn test is carried out in order to guarantee the *braking stability* of a vehicle under full braking in a turn. For that purpose, the deviation(s) of vehicle attitude or/and path is checked after complete stop. In the worst case, the vehicle gets out of a course. In consequence, it may get in an accident.

Since there usually exists a little time delay until full braking, a load transfer from a rear axle to a front axle due to braking is not as severe as powering on/off. The braking itself also reduces vehicle speed, and then it drops oversteer tendency. In addition, the unequal reductions of cornering forces at front and rear tires induced due to braking change steer characteristic. It should be considered that steering corrections are needed to bring the vehicle back on a course:

Driving speed: Limit speed.
Steering input: Fixed steering.
Procedure: Apply full braking in a turn until complete stop working or with VDC turned off.

37.3.4.4 Rollover Cornering

When a vehicle gets in an emergency situation, in general, a normal driver unconsciously applies a quick and big steering input in order to avoid an accident. In sequence, another rapid and big

countersteering input may be employed in the other direction. The inappropriate steering input makes the vehicle in a prompt turn, which causes a full load transfer from inner tires to outer tires. Depending on the rollover stability characteristic of the vehicle, roll vibration is either damped out or amplified to a rollover situation.

A rollover (cornering) test evaluates the *rollover stability* of a vehicle using specially designed test methods. For that purpose, a few test methods are so far used in practice. In this handbook, only the NHTSA rollover test method is introduced, because the others are for companies' own usages. It should be noted that this test is just an objective test operated by a steering (robot) machine to overcome its procedural difficulty. As the NHTSA's NCAP rollover rating system [48], this method consists of two processes: (1) a pretest to find a required steering input for the main test and (2) the main test to evaluate the rollover stability of a vehicle.

Using a steering (robot) machine, the pretest is carried out by gradually applying steering handwheel angle up to 270° with a ramp steering input of 13.5°/s with constant speed at 50 mph as shown in Figure 37.43. This maximum steering is kept for 2 s, and then the steering handwheel is completely released for a convenience to a driver. A certain value of steering handwheel angle causing 0.3 g of lateral acceleration is found. Then, the amplitude of a steering handwheel input to the main test is determined by multiplying the angle of 0.3 g normally by 6.5 times. 6.5 is a basic multiplier that is properly readjusted depending on a test result such as pass or failure.

For the main test, a vehicle is driven in straight ahead at a speed slightly higher than a desired entrance vehicle speed. The driver releases throttle off and simultaneously initiates a steering handwheel command described in Figure 37.44 using the steering (robot) machine. The entrance speed ranges from 35 to 45 mph and from 45 to 50 mph increased by 5 and 2.5 mph, respectively, until a termination condition is achieved. The termination is done based on either a two-wheel lift or the completion of the test through the maximum speed without the two-wheel lift. The two-wheel lift is defined as at least the 2 in. simultaneous lift of inside tires from the test surface.

For the main test shown in Figure 37.44, the first ramp steering handwheel input of 720°/s is applied up to the amplitude predetermined through the pretest. The maximum steering angle is kept until vehicle roll rate hits its band of ±1.5°/s. Afterward, the second ramp steering input of 720°/s is reversely steered up to the predetermined amplitude in the other direction. Then, the maximum steering input is kept for 3 s, and then the steering is fully released to steering center.

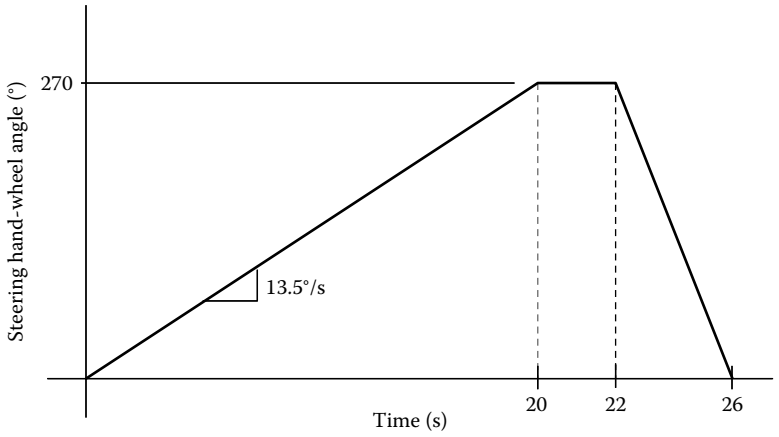


FIGURE 37.43 Steering handwheel input of a pretest of NHTSA's NCAP rollover rating system. (From Forkenbrock, G.J. et al., A demonstration of the dynamic tests developed for NHTSA's NCAP rollover rating system: Phase VIII of NHTSA's light vehicle rollover research program, NHTSA DOT HS 809 705, 2004.)

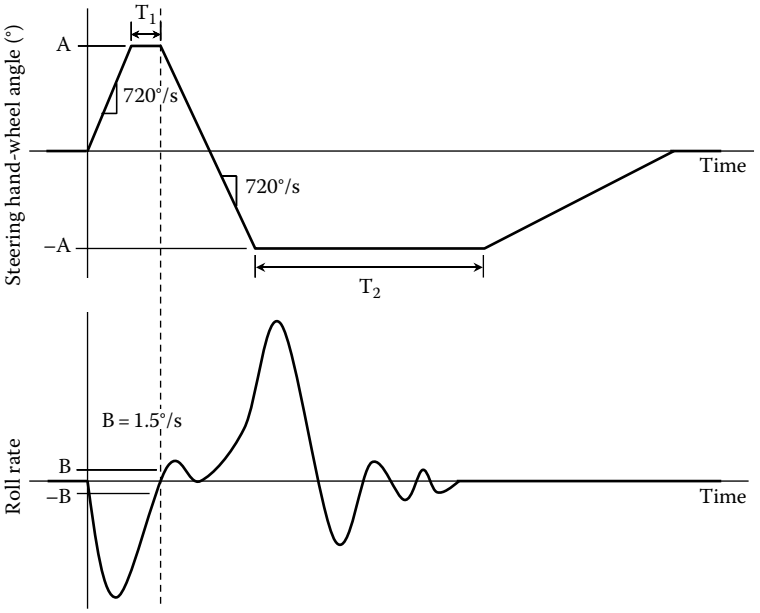


FIGURE 37.44 Steering handwheel input to and roll rate response of the main test of NHTSA's NCAP rollover rating system. (From Forkenbrock, G.J. et al., A demonstration of the dynamic tests developed for NHTSA's NCAP rollover rating system: Phase VIII of NHTSA's light vehicle rollover research program, NHTSA DOT HS 809 705, 2004.)

It should be considered that steering corrections are needed to bring the vehicle back on a course. The summary of the tests is as follows:

Driving speed: From 35 to 45 mph and from 45 to 50 mph increased by 5 and 2.5 mph, respectively.
Steering input: Series of *ramp* steerings.
Procedure: Lift throttle off and simultaneously initiate a steering handwheel command using a steering (robot) machine.

37.3.5 Riding Maneuver

For the purpose of evaluation of ride performance, a riding maneuver is generally used while driving on various public roads or highways. For a convenience, most automobile and tire makers also utilize specially designed ride roads. The ride road is just a collection of ride evaluation road segments in several tracks [40]. Sometimes, it contains some artificial segments.

In this handbook, wavy road, rough road, and impact object driving tests are introduced. Each test speed is just normal driving speed on a public road or a highway.

37.3.5.1 Wavy Road Driving

Driving on a wavy road produces the very low-frequency vibration of a sprung mass. The purpose of a wavy road driving test is to assess *primary ride* (or *body bounce*) as the vibratory characteristic of the sprung mass. Its frequency ranges up to about 5 Hz.

37.3.5.2 Rough Road Driving

When a vehicle is running on a rough road, the low-frequency vibrations of its unsprung masses are induced. A rough road driving test is generally used to evaluate *secondary ride* (or *wheel hop*) as the vibratory characteristic of the unsprung masses. Its frequency ranges from about 5 to 25 Hz.

37.3.5.3 Impact Object Driving

The purpose of an impact object driving test is to assess *impact ride* (or *impact harshness*) as the impact characteristic of a vehicle running over various impact objects. Its frequency in general ranges from about 25 to 60 Hz (at most 100 Hz). The impact objects are composed of a broken edge, a joint, a drain cover, and a cat's eye.

37.4 OBJECTIVE TEST

For various handling objective tests, ISO international standard defines the terminology of vehicle dynamics and road-holding ability [11]. In addition, it standardizes the open-loop test procedures for steady-state circular driving behavior and lateral transient response characteristics against step steering input, sinusoidal steering input, random steering input, and braking in a turn [12–16]. As a severe lane-change maneuver, the standardized test procedure of a double-lane-change test is also provided [17].

ISO international standard 2631-1 defines measurement methods for periodic, random, and transient whole-body vibration [18]. It provides the frequency-dependent weightings of sensitivity affecting health, comfort, and motion sickness due to the different axes of vibration. It plays a role to provide a basic guidance to evaluate ride quality. Jang and Griffin studied the effect of phase, frequency, magnitude, and posture on discomfort associated with differential vertical vibration at the seat and feet [49].

Since the final judgment to assess the handling and ride performances of a vehicle is carried out only through various subjective tests, the procedure and condition of an objective test should be the same as those of a specific subjective test. Any other objective test with different test procedures and conditions may provide some other useful information, but it hardly represents the subjective test performed for subjective assessment.

In this handbook, subjective and objective tests are recommended to be simultaneously carried out. Otherwise, at least the objective test must keep the same procedure and condition of the subjective test. For handling and ride objective tests, different sensors are used.

37.4.1 Handling Test

During a handling test, *steering* (*handwheel*) *angle*, *steering* (*handwheel*) *torque*, *longitudinal* and *lateral components of vehicle velocity*, *longitudinal* and *lateral components of vehicle acceleration*, *yaw rate*, *pitch rate*, and *roll rate* are generally measured [50]. For that purpose, (a) a steering wheel sensor, (b) a speed sensor, (c) an acceleration sensor, and (d) a Gyro sensor are utilized. Supplementarily, (e) a GPS can be employed to trace vehicle path.

For example, a lane changing test shown in Figure 37.32 was employed to illustrate objective measurement variables. All measurement signals were obtained with a 100 Hz sampling rate. For the purpose of filtering the sampled data, a phaseless (zero-phase-shift) digital filter was used with its passband ranging from 0 to 5 Hz. This filter is well documented in ISO international standard [12,13,16].

37.4.1.1 Steering Wheel Sensor

In order to measure steering handwheel angle, speed, and torque, a steering wheel sensor system is mounted after removing a steering handwheel as shown in Figure 37.45. Depending on resolution requirement, the steering wheel sensor should be properly chosen. For the measurement of on-center steering, a high-precision system is required (refer to Figure 37.45a). Otherwise, a standard precision system with wide ranges of measurement is used (refer to Figure 37.45b). Figure 37.46 shows the measured data of steering handwheel angle, speed, and torque obtained during lane changing.

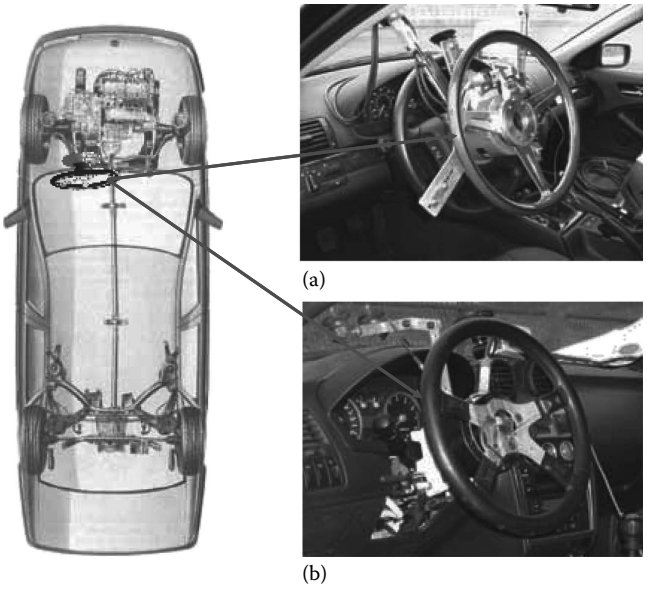


FIGURE 37.45 Steering wheel sensor. (a) High precision sensor and (b) standard precision sensor.

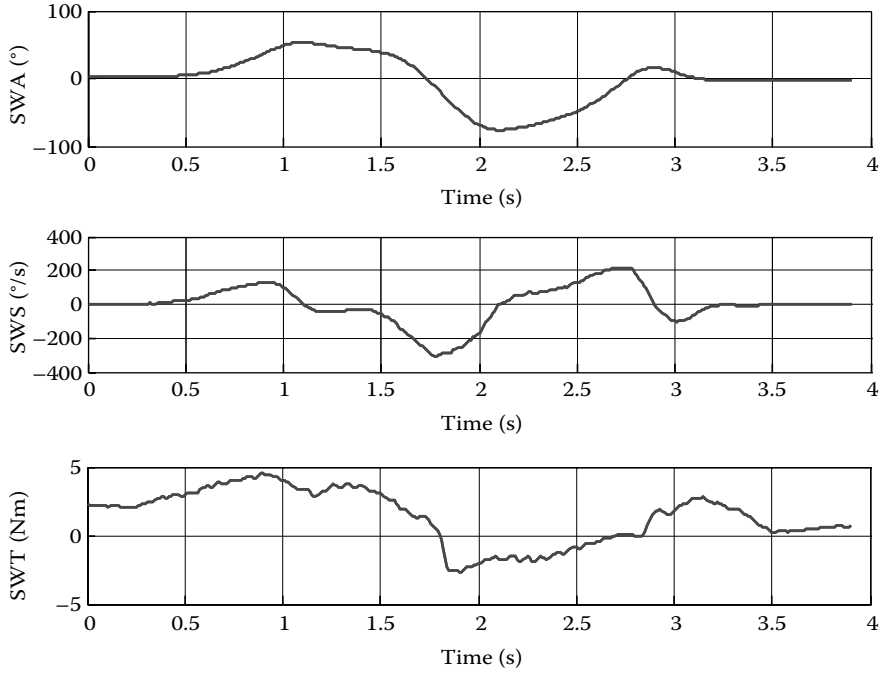


FIGURE 37.46 Steering handwheel angle, speed, and torque measured at a steering system during lane changing.

37.4.1.2 Speed Sensor

As shown in Figure 37.47, a 2-axis speed sensor is equipped at the rear position of a vehicle to gain the longitudinal and lateral components of vehicle velocity. Using this sensor, the components defined with respect to the body-fixed coordinates are measured. Figure 37.48 shows the longitudinal and lateral components measured during lane changing. While the longitudinal component is almost kept constant, the lateral component is shown various due to the directional change of the vehicle velocity.

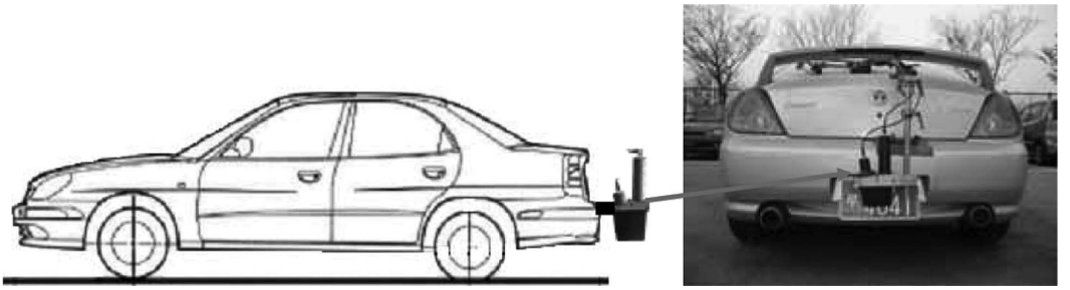


FIGURE 37.47 A speed sensor.

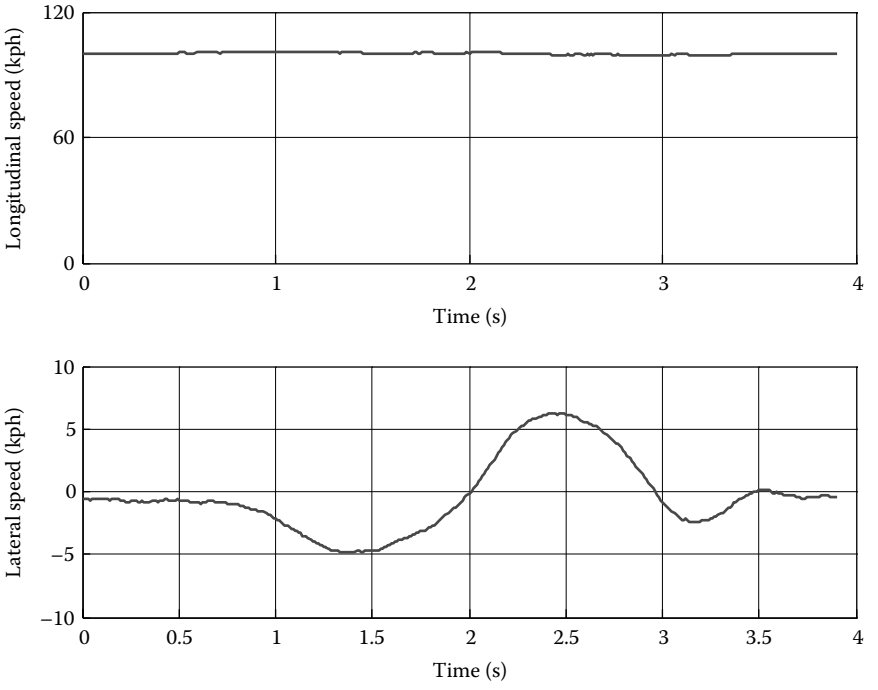


FIGURE 37.48 Longitudinal and lateral components of vehicle velocity measured at a rear position during lane changing.

Using the measured data, vehicle speed, its sideslip angle, and other vehicle motion variables are extracted. Before starting that process, the measured data should be transformed with respect to the coordinates parallel to the ground surface from the body-fixed coordinates in order to compensate roll and pitch effects. Nowadays, a speed sensor packaged together with a Gyro sensor and other sensors can be located inside of a vehicle as shown in Figure 37.49. Using this system, all the components defined with respect to both coordinates can be automatically obtained.

37.4.1.3 Acceleration Sensor

Figure 37.49 illustrates a 3-axis accelerometer installed at a rear location near a driver seat inside a vehicle. Three components of acceleration measured during lane changing are shown in Figure 37.50. Except the lateral component, the others are almost kept constant. In order to analyze the measured data regarding handling performance, in general, the lateral component is mostly utilized. In addition, the longitudinal component is used only for the purpose of analysis of power on/off or braking.



FIGURE 37.49 An accelerometer and a Gyro sensor without/with a speed sensor or/and a GPS.

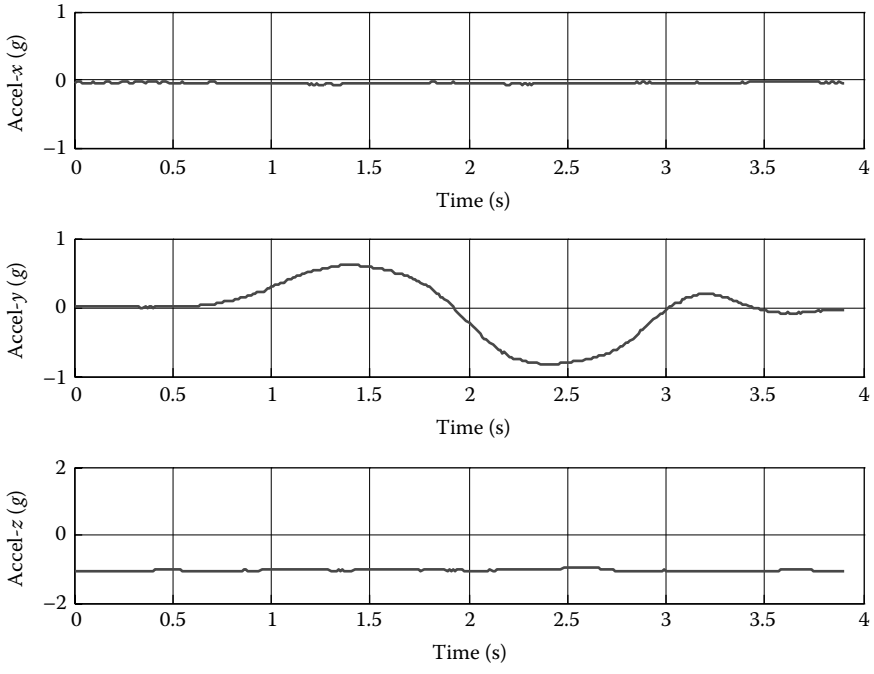


FIGURE 37.50 Longitudinal, lateral, and vertical components of acceleration measured at a rear location near a driver seat during lane changing.

Before utilizing the measured data for a vehicle dynamics analysis, the data should be transformed with respect to the coordinates parallel to the ground surface from the body-fixed coordinates to eliminate roll and pitch effects. Since these days the accelerometer is usually packaged together with a Gyro sensor, all the components defined with respect to both coordinates can be automatically gained.

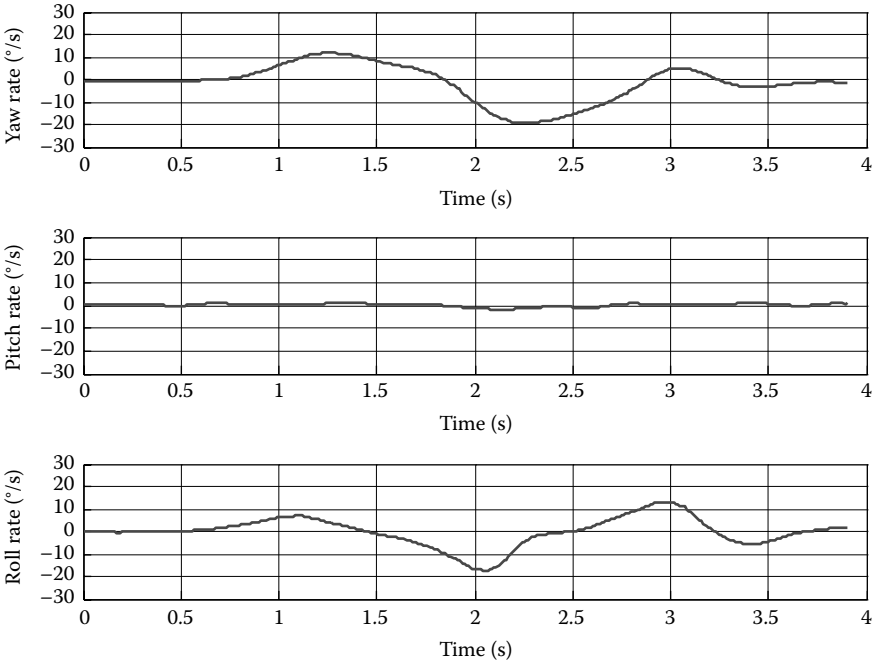


FIGURE 37.51 Yaw rate, pitch rate, and roll rate measured at a rear location near a driver seat during lane changing.

37.4.1.4 Gyro Sensor

A 3-axis Gyro sensor packaged together with an accelerometer is shown in Figure 37.49. Figure 37.51 shows three components of angular time rate measured during lane changing. Only yaw rate and roll rate are shown significant in this case. But, in general, all these data are utilized.

37.4.1.5 GPS

Sometimes a GPS is additionally used to measure vehicle path. Nowadays, the GPS packaged together with other sensors is generally employed (refer to Figure 37.49). The longitudinal and lateral coordinates of the vehicle path are given in Figure 37.52. The first and second ones present these coordinates against time, and the third one directly illustrates the vehicle path with representing its lateral coordinate against its longitudinal coordinate.

37.4.2 Ride Test

In order to analyze ride performance, six accelerometers attached to (a) a front wheel, (b) a rear wheel, (c) a vehicle floor, (d) a seat rail, (e) a seat cushion, and (f) a seat back may be needed as shown in Figure 37.53 [18]. A measurement at the seat back is more often not used because of its bad repeatability. A 3-axis accelerometer is usually used to measure the vibrations of three components of acceleration at each position. In general, the longitudinal and vertical components of the acceleration are utilized. Their frequencies range from about 0.5 to at most 100 Hz.

For example, a rectangular-shaped cleat bar on a concrete road shown in Figure 37.54 was chosen for an impact object driving test. Its height and width are 20 and 20 mm, and its length is long enough so that both left and right tires are simultaneously running over the cleat bar. For all test duration, test speed kept constant at 50 kph even while front and rear tires are passing over the cleat bar. All accelerations at six positions were measured to show the vibrations of all the positions. Their signals were measured with a 1000 Hz sampling rate and were filtered by a 100 Hz low-pass filter.

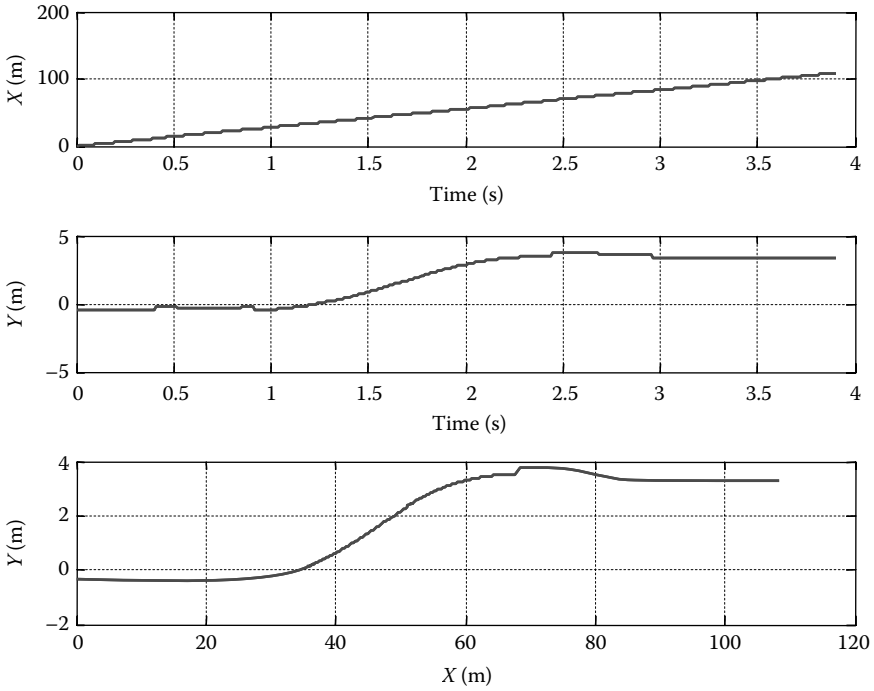


FIGURE 37.52 Longitudinal and lateral coordinates of vehicle path measured at a rear location near a driver seat during lane changing.

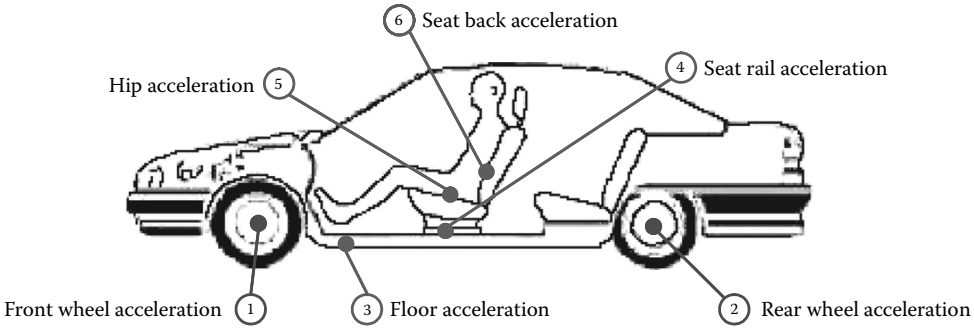


FIGURE 37.53 Measurements of accelerometers attached at five locations for ride tests.

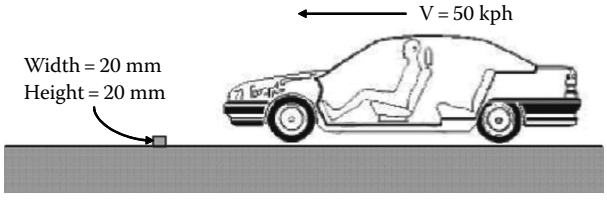


FIGURE 37.54 Driving over a rectangular-shaped cleat bar at the constant speed of 50 kph.

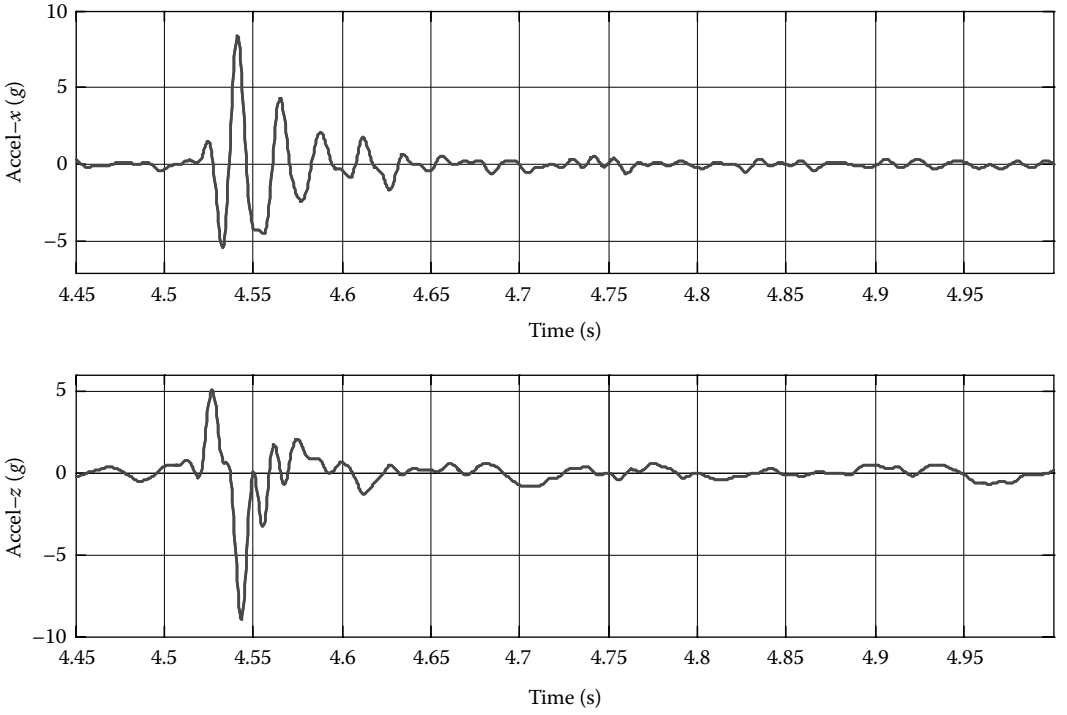


FIGURE 37.55 Longitudinal and vertical components of acceleration measured at a front knuckle during driving over a rectangular-shaped cleat bar.

37.4.2.1 Front Wheel

In order to illustrate front-wheel vibration, the longitudinal and vertical components of acceleration measured at a front knuckle during driving over the cleat bar are given in Figure 37.55. The peak value of the vertical component is bigger than that of the longitudinal component. The vertical component is damped out earlier than the longitudinal component.

37.4.2.2 Rear Wheel

Figure 37.56 shows the longitudinal and vertical components of acceleration measured at a rear knuckle to represent rear-wheel vibration. Contrary to the front-wheel vibration, the peak value of the longitudinal component is a little bigger than that of the vertical component. The longitudinal component is damped out earlier than the vertical component.

37.4.2.3 Vehicle Floor

An accelerometer is installed underneath a vehicle floor at a driver's foot position. In order to illustrate floor vibration perceived through a driver's foot, the longitudinal and vertical components of acceleration are given in Figure 37.57. While front tires are running over the cleat bar, the peak value of the vertical component is overall a little bigger in magnitude than the longitudinal component. The vertical component is damped out a little easier than the longitudinal component.

37.4.2.4 Seat Rail

Figure 37.58 shows seat-rail vibration with its longitudinal and vertical components of acceleration. The peak value of the longitudinal component is a little bigger than that of the vertical component only, while rear tires are passing over the cleat bar. After the impact of rear tires, the vertical component is not easily damped out.

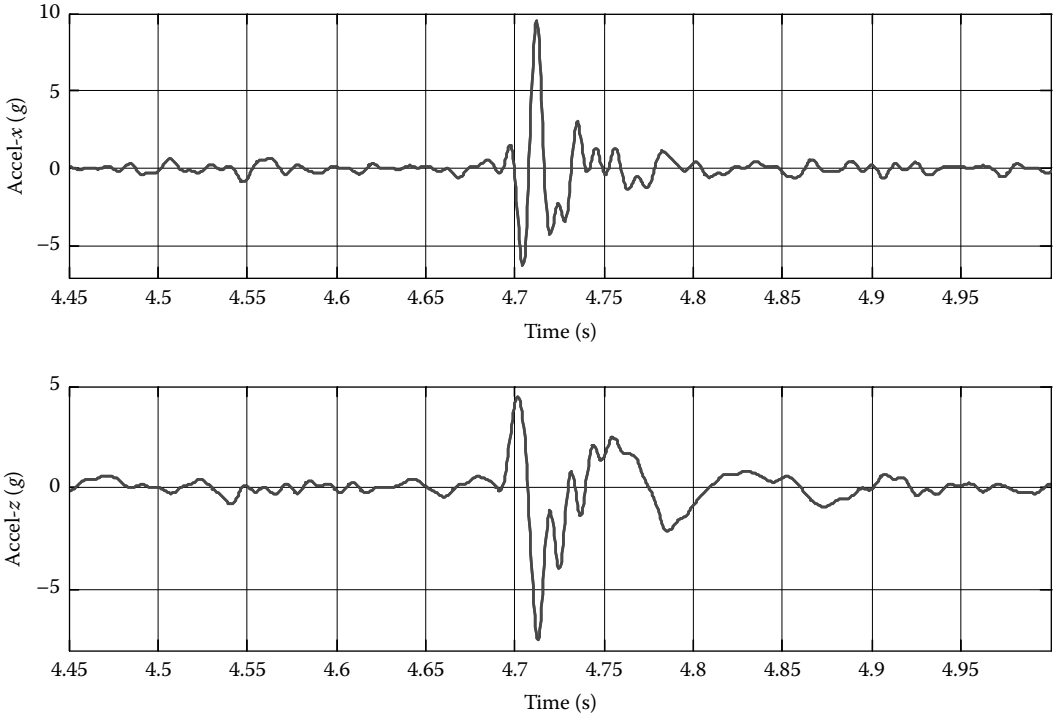


FIGURE 37.56 Longitudinal and vertical components of acceleration measured at a rear knuckle while driving over a rectangular-shaped cleat bar.

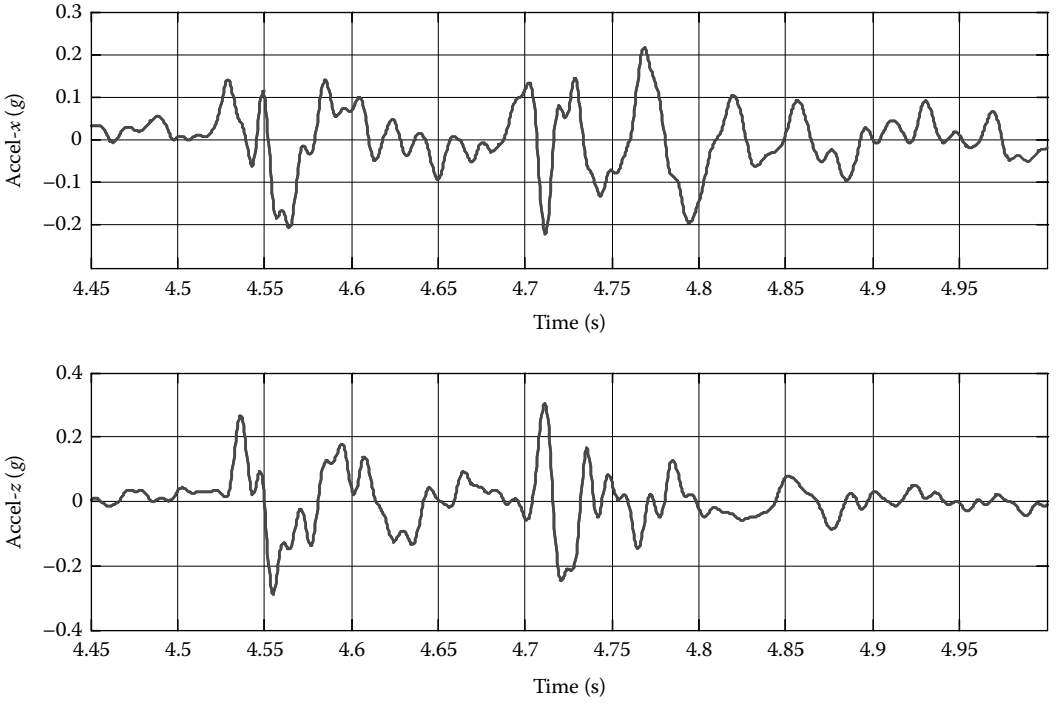


FIGURE 37.57 Longitudinal and vertical components of acceleration measured underneath vehicle floor at a driver's foot position while driving over a rectangular-shaped cleat bar.

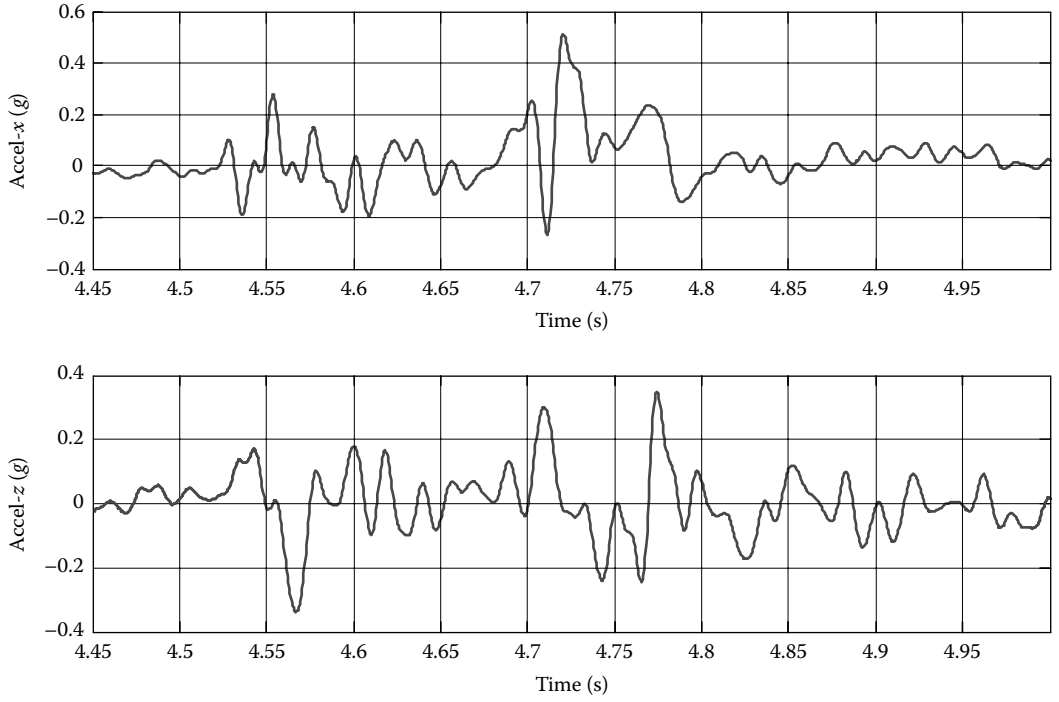


FIGURE 37.58 Longitudinal and vertical components of acceleration measured at a seat rail while driving over a rectangular-shaped cleat bar.

37.4.2.5 Seat Cushion

A seat cushion is a direct contact location, which a driver feels ride *comfortability* through. As shown in Figure 37.59, its vibration is represented by the longitudinal and vertical components of acceleration measured at the seat cushion. The peak value of the vertical component is much smaller than that of the longitudinal component during all impacts. After the impact of rear tires, both components are still vibrated for a moment.

The frequency of the vertical component is much lower than that of the longitudinal component. This phenomenon is on purpose induced due to the softness of the seat cushion.

37.4.2.6 Seat Back

A seat back is also one of the direct contact places with a driver body. Figure 37.60 illustrates the longitudinal and vertical components of acceleration measured at the seat back. The magnitude of the vertical component is much bigger than that of the longitudinal component in contrast to the seat cushion. That can be explained by the fact that the seat cushion and the seat back are designed to mostly absorb their vertical and longitudinal vibrations, respectively.

In Figures 37.58 and 37.60, the vertical component of the seat back vibration is shown bigger than that of the seat-rail vibration. This abnormal phenomenon is assumed due to the movement of a seat back sensor, which is tight to the seat back generally by fabric strings. The sensor naturally moves along the vertical direction with a driver. Therefore, the vertical component of the seat back vibration is measured dependent on a driver's leaning load on the seat back. In general, it is very difficult to get the good repeatability of the measurement of the vertical component.

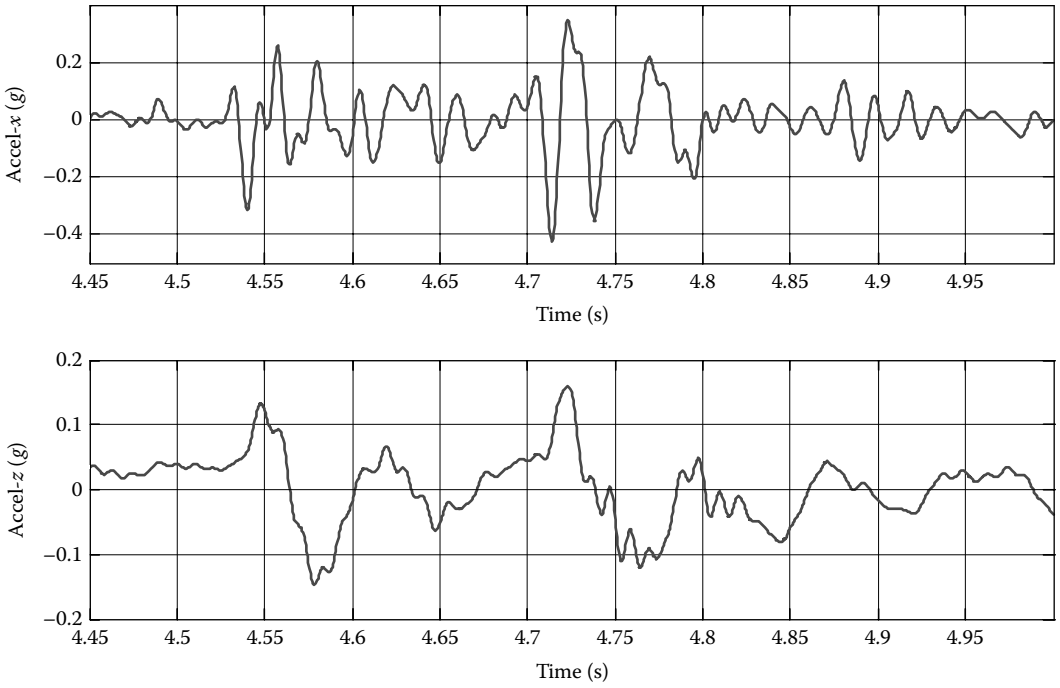


FIGURE 37.59 Longitudinal and vertical components of acceleration measured at a seat cushion while driving over a rectangular-shaped cleat bar.

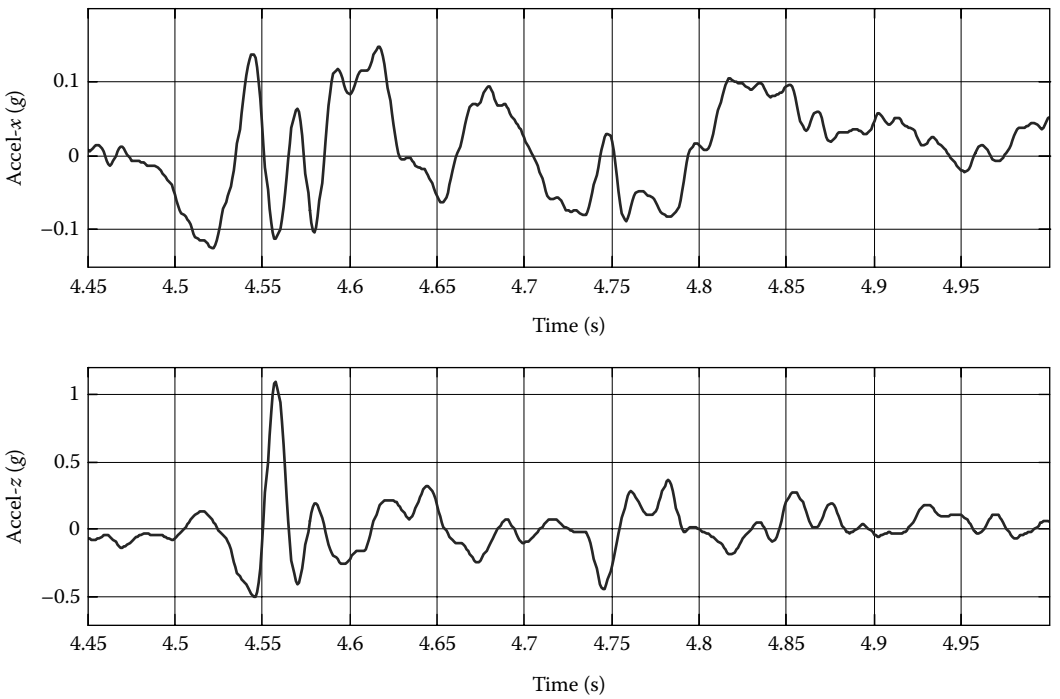


FIGURE 37.60 Longitudinal and vertical components of acceleration measured at a seat back while driving over a rectangular-shaped cleat bar.

37.5 VEHICLE DYNAMICS VARIABLES

During various subjective tests, a driver utilizes his own *subjective assessment variables* to make subjective assessment. At an instant, vehicle motion is described in terms of *vehicle motion variables*. During the tests, *objective measurement variables* are measured to illustrate the vehicle motion variables. In order to express the subjective assessment variables in mathematical formulations, *physical perception variables* are derived in terms of the vehicle motion variables. Eventually, they are extracted as functions of the objective measurement variables. In this handbook, only planar vehicle motion is presented for an example.

37.5.1 Vehicle Motion Variables

In order to describe planar vehicle motion, the center of mass, an effective front axle position, and an effective rear axle position are separately considered. As the main vehicle motion variables defined at the center of mass, *vehicle velocity V*, *turning velocity* ω, *yaw rate* $\dot{\psi}$, and *lateral acceleration* a_y are employed as shown in Figure 37.61. As the directional angle of vehicle traveling, *sideslip angle* β is defined, and *sideslip angle time rate* is denoted as $\dot{\beta}$.

For the purpose of illustrating local planar vehicle motion, *effective front steer angle* δ_{sf} and *effective rear steer angle* δ_{sr} in a steady-state turn are employed as shown in Figure 37.62. *Effective front and rear velocities* are denoted by V_f and V_r, respectively. Their directional angles are defined as *effective front slip angle* α_f and *effective rear slip angle* α_r. l, a, and b denote the distances between front and rear axles, between the center of mass and the front axle, and between the center of mass and the rear axle, respectively.

For a front axle, we have

$$V_f \sin(\delta_{sf} + \alpha_f) = V \sin\beta + a\dot{\psi}$$

Since in general $V_f \cong V$ at high speed, we get

$$\sin(\delta_{sf} + \alpha_f) \cong \sin\beta + \frac{a}{V}\dot{\psi} \tag{37.1}$$

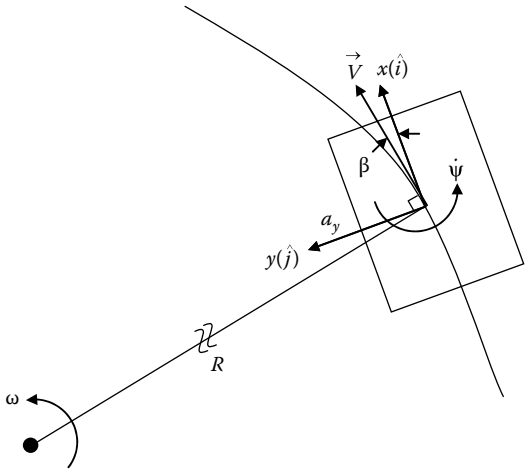


FIGURE 37.61 Vehicle motion parameters measured at the center of mass (in this situation $\beta > 0$).

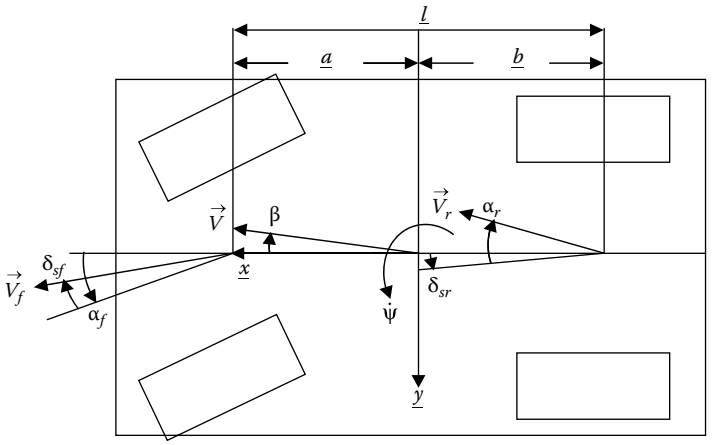


FIGURE 37.62 Effective steer and slip angles in a steady-state turn: $\delta_{sf} > 0$, $\delta_{sr} > 0$, $\beta < 0$, $\alpha_f < 0$, and $\alpha_r < 0$.

Similar to the front axle, for a rear axle, we also have

$$V_r \sin(\delta_{sr} + \alpha_r) = V \sin\beta - b\dot{\psi}$$

Since for most cases $V_r \cong V$ at high speed, we also get

$$\sin(\delta_{sr} + \alpha_r) \cong \sin\beta - \frac{b}{V}\dot{\psi} \tag{37.2}$$

Equations 37.1 and 37.2 yield

$$\sin(\delta_{sf} + \alpha_f) - \sin(\delta_{sr} + \alpha_r) = \frac{1}{V}\dot{\psi} \tag{37.3}$$

If R is defined as turning radius, for a steady-state turn, we have

$$V = R\dot{\psi} \tag{37.4}$$

Substitution of Equation 37.4 in Equation 37.3 becomes

$$\sin(\delta_{sf} + \alpha_f) - \sin(\delta_{sr} + \alpha_r) = \frac{l}{R} \tag{37.5}$$

Then,

$$\delta_{sf} = \sin^{-1}\left[\frac{l}{R} + \sin(\delta_{sr} + \alpha_r)\right] - \alpha_f \tag{37.6}$$

The partial derivative of Equation 37.5 yields

$$\frac{\partial \sin(\delta_{sf} + \alpha_f)}{\partial a_y} = \frac{\partial \sin(\delta_{sr} + \alpha_r)}{\partial a_y}$$

and

$$\cos(\delta_{sf}+\alpha_f)\frac{\partial(\delta_{sf}+\alpha_f)}{\partial a_y}=\cos(\delta_{sr}+\alpha_r)\frac{\partial(\delta_{sr}+\alpha_r)}{\partial a_y}$$

Then, we have

$$\frac{\partial\delta_{sf}}{\partial a_y}=\frac{\cos(\delta_{sr}+\alpha_r)}{\cos(\delta_{sf}+\alpha_f)}\frac{\partial(\delta_{sr}+\alpha_r)}{\partial a_y}-\frac{\partial\alpha_f}{\partial a_y}$$

For a small cornering activity, in case $\delta_{sr}\cong 0$, we get

$$\frac{\partial\delta_{sf}}{\partial a_y}\cong\frac{\partial(-\alpha_f+\alpha_r)}{\partial a_y} \tag{37.7}$$

For a steady-state turn as shown in Figure 37.63, from Equation 37.2, we have

$$\sin(\delta_{sr}+\alpha_r)=\sin\beta-\frac{b}{R} \tag{37.8}$$

The partial derivative of Equation 37.8 gives

$$\frac{\partial\sin(\delta_{sr}+\alpha_r)}{\partial a_y}=\frac{\partial\sin\beta}{\partial a_y}$$

and

$$\cos(\delta_{sr}+\alpha_r)\frac{\partial(\delta_{sr}+\alpha_r)}{\partial a_y}=\cos\beta\frac{\partial\beta}{\partial a_y}$$

Then, we get

$$\frac{\partial\beta}{\partial a_y}=\frac{\cos(\delta_{sr}+\alpha_r)}{\cos\beta}\frac{\partial(\delta_r+\alpha_r)}{\partial a_y} \tag{37.9}$$

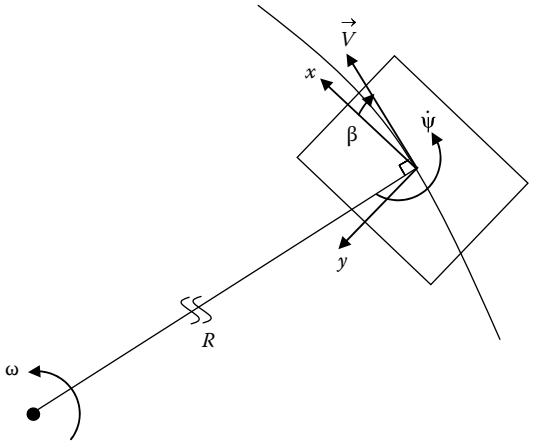


FIGURE 37.63 Steady-state turn (in this situation $\beta < 0$).

For a small cornering activity, in case $\delta_{sr} \cong 0$, we have

$$\frac{\partial \beta}{\partial a_y} = \frac{\partial \alpha_r}{\partial a_y} \tag{37.10}$$

Vehicle velocity at the center of mass is written as

$$\vec{V} = V_x \hat{i} + V_y \hat{j} \tag{37.11}$$

where

$\hat{i}$ and $\hat{j}$ are the longitudinal and lateral directional vectors fixed at the center of mass of a vehicle, respectively, as shown in Figure 37.61

V_x and V_y are the longitudinal and lateral components of V along $\hat{i}$ and $\hat{j}$, respectively

As the directional angle of vehicle moving, sideslip angle β is defined as

$$\beta \equiv \tan^{-1} \frac{V_y}{V_x} \tag{37.12}$$

Then, we get

$$V_x = V \cos\beta \tag{37.13a}$$

and

$$V_y = V \sin\beta \tag{37.13b}$$

The differentiation of Equation 37.11 yields

$$\frac{d\vec{V}}{dt} = \frac{dV_x}{dt}\hat{i} + V_x \frac{d\hat{i}}{dt} + \frac{dV_y}{dt}\hat{j} + V_y \frac{d\hat{j}}{dt} \tag{37.14}$$

The differentiations of Equations 37.13a and b also produce

$$\frac{dV_x}{dt} = \dot{V}\cos\beta - V\dot{\beta}\sin\beta \tag{37.15a}$$

and

$$\frac{dV_y}{dt} = \dot{V}\sin\beta + V\dot{\beta}\cos\beta \tag{37.15b}$$

If $\hat{k}$ denotes the directional vector perpendicular to both $\hat{i}$ and $\hat{j}$, we get

$$\frac{d\hat{i}}{dt} = \dot{\psi}\hat{k} \times \hat{i} = \dot{\psi}\hat{j} \tag{37.16a}$$

and

$$\frac{d\hat{j}}{dt} = \dot{\psi}\hat{k} \times \hat{j} = -\dot{\psi}\hat{i} \tag{37.16b}$$

Substitution of Equations 37.15a and 37.16b in Equation 37.14 yields

$$\frac{d\vec{V}}{dt} = (\dot{V}\cos\beta - V\dot{\beta}\sin\beta)\hat{i} + V\dot{\psi}\cos\beta\hat{j} + (\dot{V}\sin\beta - V\dot{\beta}\cos\beta)\hat{j} - V\dot{\psi}\sin\beta\hat{i}$$

Then, we get

$$\frac{d\vec{V}}{dt} = [\dot{V}\cos\beta - V(\dot{\psi}+\dot{\beta})\sin\beta]\hat{i} + [\dot{V}\sin\beta - V(\dot{\psi}+\dot{\beta})\cos\beta]\hat{j} \tag{37.17}$$

Now, we define $\dot{V}_x$ and $\dot{V}_y$ as

$$\dot{V}_x \equiv \dot{V}\cos\beta \tag{37.18a}$$

and

$$\dot{V}_y \equiv \dot{V}\sin\beta \tag{37.18b}$$

Using Equations 37.13a,b, 37.18a and b, Equation 37.17 is written as

$$\frac{d\vec{V}}{dt} = [\dot{V}_x - V_y(\dot{\psi}+\dot{\beta})]\hat{i} + [\dot{V}_y + V_x(\dot{\psi}+\dot{\beta})]\hat{j} \tag{37.19}$$

If a_x and a_y denote the longitudinal and lateral components of acceleration along $\hat{i}$ and $\hat{j}$, respectively, we get

$$a_x = \dot{V}_x - V_y(\dot{\psi}+\dot{\beta}) \tag{37.20a}$$

and

$$a_y = \dot{V}_y + V_x(\dot{\psi}+\dot{\beta}) \tag{37.20b}$$

Now, we get turning velocity as

$$\omega = \dot{\psi} + \dot{\beta} \tag{37.21}$$

In a steady-state turn shown in Figure 37.63, there exists sideslip outward, in which β becomes negative, that is, $\beta < 0$ in Equations 37.1 through 37.21.

37.5.2 Physical Perception Variables

A professional test engineer is observed to carry out his or her subjective assessment using his or her physical perception variables. For example, the physical perception variables only in planar vehicle motion are illustrated in this handbook. First of all, *steering handwheel angle* δ_{sw} and *steering handwheel torque* T_{sw} are perceived through a steering handwheel. A driver also senses *vehicle*

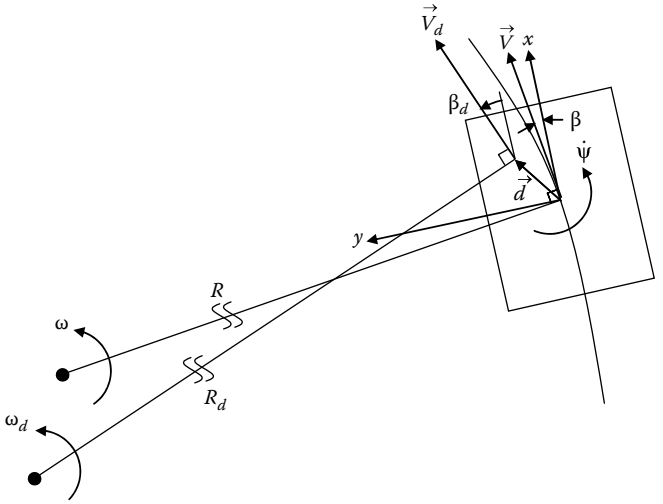


FIGURE 37.64 Driver perception parameters (in this situation $\beta > 0$ and $\beta_d > 0$).

speed V_{xd}, *turning curvature* κ_d, *turning velocity* ω_d, *nose angle* φ_d, and *nose angle time rate* $\dot{\varphi}_d$ through the reactions of vehicle path and attitude with respect to his or her seat [28,43].

As shown in Figure 37.64, vehicle velocity measured at a driver seat is symbolized by $\vec{V}_d$. A distance vector from the center of mass to a driver seat is denoted as $\vec{d}$. If the longitudinal and lateral components of $\vec{V}_d$ and $\vec{d}$ are defined along $\hat{i}$ and $\hat{j}$, they are denoted as V_{xd}, V_{yd}, d_x, and d_y, respectively. Now, we have

$$\vec{V}_d = V_x\hat{i} + V_y\hat{j} + \dot{\psi}\hat{k} \times (d_x\hat{i} + d_y\hat{j})$$

Subscript "*d*" stands for "*measured at a driver seat*" or "*with respect to a driver seat.*"

Then, we have

$$\vec{V}_d = (V_x - d_y\dot{\psi})\hat{i} + (V_y + d_x\dot{\psi})\hat{j} \tag{37.22}$$

It is also simplified as

$$\vec{V}_d = V_{xd}\,\hat{i} + V_{yd}\,\hat{j} \tag{37.23}$$

Then, Equations 37.22 and 37.23 yield

$$V_{xd} = V_x - d_y\dot{\psi} \tag{37.24a}$$

and

$$V_{yd} = V_y + d_x\dot{\psi} \tag{37.24b}$$

Substitution of Equations 37.13a and b in Equation 37.22 yields

$$\vec{V}_d = (V\cos\beta - d_y\dot{\psi})\hat{i} + (V\sin\beta - d_x\dot{\psi})\hat{j}$$

Its differentiation yields

$$\frac{d\vec{V}_d}{dt} = (\dot{V}\cos\beta - V\dot{\beta}\sin\beta - d_y\ddot{\psi})\hat{i} + (V\cos\beta - d_y\dot{\psi})\frac{d\hat{i}}{dt} + (\dot{V}\sin\beta - V\dot{\beta}\cos\beta - d_x\ddot{\psi})\hat{j} + (V\sin\beta - d_x\dot{\psi})\frac{d\hat{j}}{dt}$$

Using Equations 37.16a and b, we get

$$\frac{d\vec{V}_d}{dt} = (\dot{V}\cos\beta - V\dot{\beta}\sin\beta - d_y\ddot{\psi})\hat{i} + (V\cos\beta - d_y\dot{\psi})\dot{\psi}\hat{j} + (\dot{V}\sin\beta - V\dot{\beta}\cos\beta - d_x\ddot{\psi})\hat{j} + (V\sin\beta - d_x\dot{\psi})\dot{\psi}\hat{i}$$

Then, we have

$$\frac{d\vec{V}_d}{dt} = [\dot{V}\cos\beta - V(\dot{\psi}+\dot{\beta})\sin\beta - d_x\dot{\psi}^2 - d_y\ddot{\psi}]\hat{i} + [\dot{V}\sin\beta + V(\dot{\psi}+\dot{\beta})\cos\beta - d_y\dot{\psi}^2 + d_x\ddot{\psi}]\hat{j} \quad (37.25)$$

Substitution of Equations 37.13a,b, 37.18a and b in Equation 37.25 yields

$$\frac{d\vec{V}_d}{dt} = [\dot{V}_x - V_y(\dot{\psi}+\dot{\beta}) - d_x\dot{\psi}^2 - d_y\ddot{\psi}]\hat{i} + [\dot{V}_y + V_x(\dot{\psi}+\dot{\beta}) - d_y\dot{\psi}^2 + d_x\ddot{\psi}]\hat{j} \quad (37.26)$$

If a_{xd} and a_{xd} denote the longitudinal and lateral components of acceleration measured at a driver seat along $\hat{i}$ and $\hat{j}$, respectively, Equation 37.26 yields

$$a_{xd} = \dot{V}_x - V_y(\dot{\psi}+\dot{\beta}) - d_x\dot{\psi}^2 - d_y\ddot{\psi}$$

and

$$a_{yd} = \dot{V}_y - V_x(\dot{\psi}+\dot{\beta}) - d_y\dot{\psi}^2 - d_x\ddot{\psi}$$

Usage of Equations 37.20a and b gives

$$a_{xd} = a_x - d_x\dot{\psi}^2 - d_y\ddot{\psi} \tag{37.27a}$$

and

$$a_{yd} = a_y - d_y\dot{\psi}^2 + d_x\ddot{\psi} \tag{37.27b}$$

With known V_{xd} and V_{yd}, we get vehicle speed perceived at a driver seat as

$$V_d = \sqrt{V_{xd}^2 + V_{yd}^2} \tag{37.28}$$

If β_d is defined as the directional angle of vehicle moving measured at a driver seat, we get

$$\beta_d \equiv \tan^{-1}\frac{V_{yd}}{V_{xd}} \tag{37.29}$$

Then, we have

$$V_{xd} = V_d\cos\beta_d \tag{37.30a}$$

and

$$V_{yd} = V_d\sin\beta_d \tag{37.30b}$$

From Equation 37.23, we get

$$\frac{d\vec{V}_d}{dt} = \frac{dV_{xd}}{dt}\hat{i} + V_{xd}\frac{d\hat{i}}{dt} + \frac{dV_{yd}}{dt}\hat{j} + V_{yd}\frac{d\hat{j}}{dt} \tag{37.31}$$

From Equations 37.30a and b, we have

$$\frac{dV_{yd}}{dt} = \dot{V}_d\cos\beta_d + V_d\dot{\beta}_d\sin\beta_d \tag{37.32a}$$

and

$$\frac{dV_{xd}}{dt} = \dot{V}_d\sin\beta_d + V_d\dot{\beta}_d\cos\beta_d \tag{37.32b}$$

Substitution of Equations 37.16a,b, 37.32a and b in Equation 37.31 yields

$$\begin{aligned}\frac{d\vec{V}_d}{dt} &= (\dot{V}_d\cos\beta_d - V_d\dot{\beta}_d\cos\beta_d)\hat{i} + V_d\dot{\psi}\cos\beta_d\hat{j} \\ &\quad + (\dot{V}_d\sin\beta_d + V_d\dot{\beta}_d\cos\beta_d)\hat{j} - V_d\dot{\psi}\sin\beta_d\hat{i}\end{aligned}$$

Then,

$$\frac{d\vec{V}_d}{dt} = [\dot{V}_d \cos\beta_d - V_d(\dot{\psi} + \dot{\beta}_d)\sin\beta_d]\hat{i} + [\dot{V}_d \sin\beta_d + V_d(\dot{\psi} + \dot{\beta}_d)\cos\beta_d]\hat{j} \tag{37.33}$$

We define $\dot{V}_{xd}$ and $\dot{V}_{yd}$ as

$$\dot{V}_{xd} \equiv \dot{V}_d \cos\beta_d \tag{37.34a}$$

and

$$\dot{V}_{yd} \equiv \dot{V}_d \sin\beta_d \tag{37.34b}$$

Substitution of Equations 37.30a,b, 37.34a and b in Equation 37.33 yields

$$\frac{d\vec{V}_d}{dt} = [\dot{V}_{xd} - V_{yd}(\dot{\psi} + \dot{\beta}_d)]\hat{i} + [\dot{V}_{yd} - V_{xd}(\dot{\psi} + \dot{\beta}_d)]\hat{j} \tag{37.35}$$

Equation 37.35 gives

$$a_{xd} = \dot{V}_{xd} - V_{yd}(\dot{\psi} + \dot{\beta}_d) \tag{37.36a}$$

and

$$a_{yd} = \dot{V}_{yd} - V_{xd}(\dot{\psi} + \dot{\beta}_d) \tag{37.36b}$$

If turning velocity that a driver feels at his seat is denoted as ω_d, Equation 37.36b is expressed as

$$a_{yd} = \dot{V}_{yd} + V_{xd}\omega_d \tag{37.37}$$

Known a_{yd}, ω_d is computed as

$$\omega_d = \frac{a_{yd} - \dot{V}_{yd}}{V_{xd}} \tag{37.38}$$

Equations 37.36b and 37.37 also give

$$\omega_d = \dot{\psi} + \dot{\beta}_d \tag{37.39}$$

Now, turning curvature that a driver feels at his seat is determined as

$$\kappa_d = \frac{\omega_d}{V_{xd}} \tag{37.40}$$

where R_d is turning radius measured at the driver seat and $\kappa_d = 1/R_d$.

In order to describe vehicle attitude that a driver feels at his seat, nose angle measured at the driver seat is denoted as φ_d. Then, it is obtained as

$$\varphi_d = -\beta_d \tag{37.41}$$

Usage of Equations 37.41 and 37.29 yields

$$\varphi d = -\tan^{-1}\frac{V_{yd}}{V_{xd}} \tag{37.42}$$

Since for nose angle time rate we have $\dot{\varphi}_d = \dot{\beta}_d, \dot{\varphi}_d$ is determined using Equation 37.39 as

$$\dot{\varphi}_d = \dot{\psi} - \omega_d \tag{37.43}$$

It is noted that κ_d and ω_d represent vehicle path that a driver feels while driving in a turn. In addition, φ_d and $\dot{\varphi}_d$ represent vehicle attitude that a driver feels.

37.5.3 Objective Measurement Variables

While a steering wheel sensor is equipped at a steering system, a velocity sensor, a Gyro sensor, and an accelerometer are usually located away from the driver seat (refer to Figures 37.45, 37.47, and 37.49). Through the steering wheel sensor, *steering handwheel angle* δ_{sw} and *steering handwheel torque* T_{sw} are measured as shown in Figure 37.65. The velocity sensor measures *longitudinal velocity* V_{xm} and *lateral velocity* V_{ym}. Using the Gyro sensor and accelerometer, *yaw rate* $\dot{\psi}_m$ and *lateral acceleration* a_{ym} are measured, respectively.

Distance vectors from the velocity sensor and the Gyro sensor to the driver seat are denoted as $\vec{c}$ and $\vec{e}$, respectively. If their longitudinal and lateral components along $\hat{i}$ and $\hat{j}$ are also symbolized as c_x, c_y, e_x, and e_y, respectively, then we have

$$\vec{c} = c_x\hat{i} + c_y\hat{j} \quad \text{for a velocity sensor} \tag{37.44a}$$

$$\vec{e} = e_x\hat{i} + e_y\hat{j} \quad \text{for a Gyro sensor and an accelerometer} \tag{37.44b}$$

$\vec{V}_m$ denotes a measured velocity vector, and its longitudinal and lateral components along $\hat{i}$ and $\hat{j}$ are denoted as V_{xm} and V_{ym}, respectively. Then, $\vec{V}_m$ is written as

$$\vec{V}_m = V_{xm}\hat{i} + V_{ym}\hat{j} \tag{37.45}$$

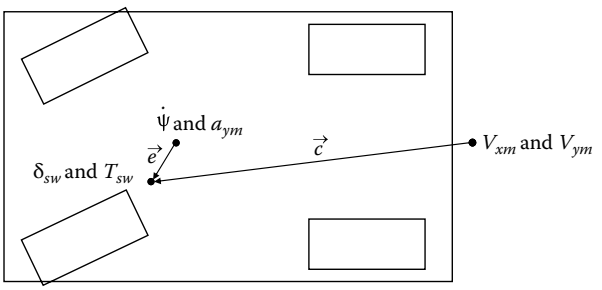


FIGURE 37.65 Objective measurement parameters.

Using Equations 37.24a and b, vehicle velocity felt by a driver can be computed as

$$V_{xd} = V_{xm} - c_y \dot{\psi}_m \tag{37.46a}$$

and

$$V_{yd} = V_{ym} - c_x \dot{\psi}_m \tag{37.46b}$$

Based on an assumption that a vehicle chassis behaves almost rigid against a low-frequency steering input, we have

$$\dot{\psi} = \dot{\psi}_m \tag{37.47}$$

If a_{ym} denotes measured lateral acceleration, usage of Equation 37.27b yields

$$a_{yd} = a_{ym} - e_y \dot{\psi}_m^2 + e_x \ddot{\psi}_m \tag{37.48}$$

where $\ddot{\psi}_m$ is electrically or numerically computed by differentiating $\dot{\psi}_m$.

As physical perception variables, *steering handwheel angle* δ_{sw} and *steering handwheel torque* T_{sw} are directly served. Using Equations 37.29, 37.34b, 37.38, 37.40, 37.42, and 37.43, other objective measurement variables are converted to additional physical perception variables such as *turning curvature* k_d, *turning velocity* ω_d, *nose* angle φ_d, and *nose angle time rate* $\dot{\varphi}_d$.

37.6 VEHICLE DYNAMICS ANALYSIS

As already explained, subjective assessment on handling performance is composed of more than twenty characteristics (refer to Figure 37.1, Tables 37.3 through 37.5). Among them, nowadays, only steady-state steer characteristic is written in forms of mathematical formulation based on a bicycle model of conventional vehicle dynamics [1–7]. The conventional vehicle dynamics cannot explain transient steer characteristics [26–28]. That's why all other characteristics still cannot be described in mathematical formulations. In this handbook, for transient steer characteristics, a new approach of vehicle dynamics analysis is introduced [28].

In order to evaluate the steer characteristic, (a) *steady-state steer*, (b) *transient understeer*, (c) *transient oversteer*, and (d) *power-off reaction steer* have been subjectively assessed so far. The steady-state steer represents a balance between front and rear axle grips during steady-state cornering [6–8]. It has been assessed for a long time through both dry and wet handling tests. However, these days, it is no longer used for a dry handling test; instead, it is still utilized for a wet handling test. The transient understeer, transient oversteer, and power-off reaction steer are mainly evaluated for a dry handling test [40–43].

The understeer or oversteer is perceptually defined in order to evaluate the grip capability of a front or a rear axle, respectively, against the required grip to get both righteous vehicle path and attitude in a transient turn. Power-off reaction steer subjectively measures the lack of rear axle grip relative to front axle grip caused due to a sudden load shift from the rear axle to the front axle while powering off immediately just after full powering on with fixed steering in a turn. All the steer characteristics are assessed through various subjective tests of (a) *lane changing*, (b) *severe lane changing*, (c) *circuit driving*, and (d) *powering on/off in a turn*.

37.6.1 FORMULATION

For the purpose of closely describing a driver's perception on steer characteristics, physical perception variables are employed. Furthermore, the steer characteristics are objectified using (a) *steer gradient* (*SG*), (b) *understeer coefficient* (*US*), (c) *oversteer coefficient* (*OS*), and (d) *power-off reaction steer coefficient*.

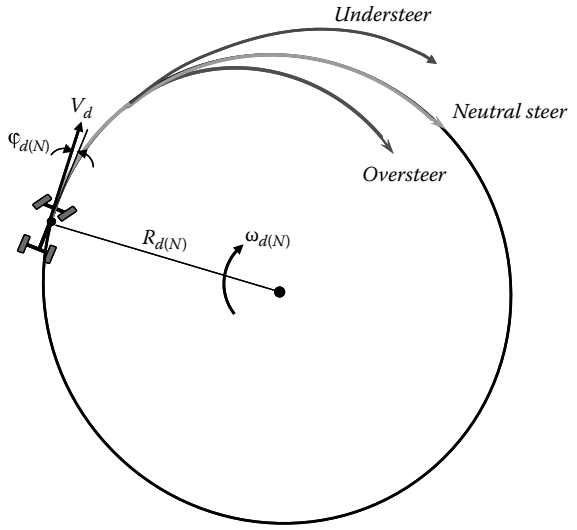


FIGURE 37.66 Steady-state neutral steer during steady-state cornering at a steering pad.

37.6.1.1 Steer Gradient

Figure 37.66 shows a steering pad test with fixed steering at constant speed for the evaluation of the traditional steady-state steer characteristic of a vehicle [6–8]. Here, (*steady-state*) *neutral steer* represents just steady-state turning characteristics with a constant turning radius and fixed steering. Once vehicle speed is very gently increased, the vehicle no longer keeps the constant radius turn due to the increase of a centrifugal force, that is, the neutral steer no longer lasts. In this situation, a balance between front and rear axle grips no longer remains. The lack of grip at one axle compared to the other produces more slip at its axle, which alters the turning radius.

In order to keep the same vehicle path, a steering handwheel should be properly adjusted. When a vehicle needs more steering angle, it is defined that the vehicle behaves *steady-state understeer*. If less steering angle is needed, on the other hand, it is defined for the vehicle to behave *steady-state oversteer*. As an objective index to evaluate the steady-state steer characteristic, *SG* is defined as a gradient of steering handwheel angle δ_{sw} with respect to lateral acceleration a_y. It is written by the SAEJ670e as [51]

$$SG \equiv \frac{\partial \delta_{sw}}{\partial a_y} \begin{cases} > 0 & \text{for steady-state understeer} \\ = 0 & \text{for steady-state neutral steer} \\ < 0 & \text{for steady-state oversteer} \end{cases} \tag{37.49}$$

37.6.1.2 Understeer Coefficient

Transient understeer characteristic is observed to be subjectively evaluated by vehicle path [28,43]. In order to represent the vehicle path as a function of *physical perception variables*, *turning curvature* κ_d and *turning velocity* ω_d are defined at a driver seat (refer to Figure 37.67). They quantitatively describe a driver's perceptual turning curvature and turning fastness, respectively. As shown in Figure 37.67, turning radius measured at the driver seat is denoted as R_d. The turning curvature is determined as a function of the turning radius, that is, $\kappa_d = 1/R_d$.

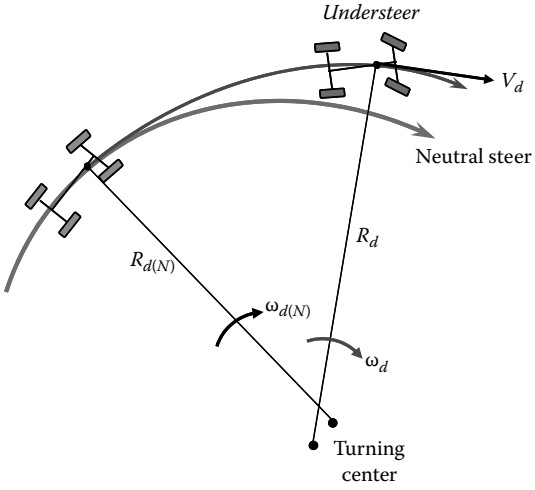


FIGURE 37.67 Turning radius and turning velocity representing transient understeer.

Using lateral acceleration a_{yd} and longitudinal velocity V_{xd} measured at the driver's seat, ω_d is computed using Equation 37.38 as

$$\omega_d = \frac{a_{yd} - \dot{V}_{yd}}{V_{xd}} \tag{37.38}$$

From Equation 37.40, we get

$$\kappa_d = \frac{\omega_d}{V_{xd}} \tag{37.40}$$

If $\kappa_{d(N)}$ and $\omega_{d(N)}$ are defined to represent κ_d and ω_d of steady-state neutral steer (refer to Figure 37.66), the transient understeer is formulated as

$$\Delta\kappa_d = sign(\Delta t_k)\left|\kappa_{d(N)} - \kappa_d\right| > 0 \quad \text{for transient understeer} \tag{37.50}$$

and

$$\Delta\omega_d = sign(\Delta t_\omega)\left|\omega_{d(N)} - \omega_d\right| > 0 \quad \text{for transient understeer} \tag{37.51}$$

where Δt_κ and Δt_ω are the time delays of κ_d and ω_d against $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively. In case κ_d and ω_d are following in behind $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively, Δt_κ and Δt_ω are positive. Therefore, the signs of $\Delta\kappa_d$ and $\Delta\omega_d$ just represent the time delays. Their magnitudes stand for the amounts of the transient understeer. Since the transient understeer is also dependent on driver's perceptual steering handwheel angle, κ_d and $\kappa_{d(N)}$ of Equation 37.50 should be measured at the same steering handwheel angle. Furthermore, the same rule is also applied to Equation 37.51.

In order to describe the transient understeer in a formulation, *US* is defined as

$$US \equiv w_{us1}\Delta\kappa_d + w_{us2}\Delta\omega_d > 0 \tag{37.52}$$

where w_{us1} and w_{us2} are understeer weighing factors, and they are dependent on each driver, vehicle, and test condition.

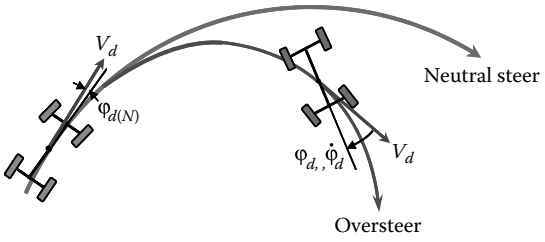


FIGURE 37.68 Nose angle and nose angle time rate representing transient oversteer.

37.6.1.3 Oversteer Coefficient

Contrary to the transient understeer, vehicle attitude is observed to subjectively determine *transient oversteer* characteristics [28,43]. For the purpose of designating the vehicle attitude, nose angle φ_d and its time rate $\dot{\varphi}_d$ defined at the driver seat are used as shown in Figure 37.68. They quantify driver's perceptual vehicle nose-in attitude and nose-in variational quickness. Then, it is determined as

$$\varphi_d = -\tan^{-1}\frac{V_{yd}}{V_{xd}}\begin{cases} > 0 & \text{for nose in} \\ = 0 & \text{for neutral nose} \\ < 0 & \text{for nose out} \end{cases} \tag{37.42}$$

where V_{xd} and V_{yd} are the longitudinal and lateral components of V_d. As the quickness of nose-in changes, the time rate of φ_d is obtained from Equation 37.43 as

$$\dot{\varphi}_d = \dot{\psi} - \omega_d \tag{37.43}$$

Similar to the transient understeer, the transient oversteer is determined as

$$\Delta\varphi_d = sign(\varphi_d)(\varphi_d - \varphi_{d(N)}) > 0 \quad \text{for transient oversteer} \tag{37.53}$$

where $\varphi_{d(N)}$ refers to φ_d of steady-state neutral steer (refer to Figure 37.66). Since $\dot{\varphi}_d$ representing the neutral steer is believed to be zero, we have

$$\Delta\dot{\varphi}_d = sign(\dot{\varphi}_d)\dot{\varphi}_d > 0 \quad \text{for transient oversteer} \tag{37.54}$$

For the purpose of formulation of the transient oversteer, OS is defined as

$$OS \equiv w_{os1}\Delta\omega_d + w_{os2}\Delta\dot{\varphi}_d > 0 \tag{37.55}$$

where w_{os1} and w_{os2} are oversteer weighing factors, and they also depend on each driver, vehicle, and test condition.

37.6.1.4 Power-Off Reaction Steer Coefficient

During powering off immediately just after full powering on with fixed steering, the reaction of a vehicle mostly affects the impulsive change of vehicle attitude [28,43]. A driver subjectively assesses power-off reaction steer characteristics based on his or her perception of instantaneous

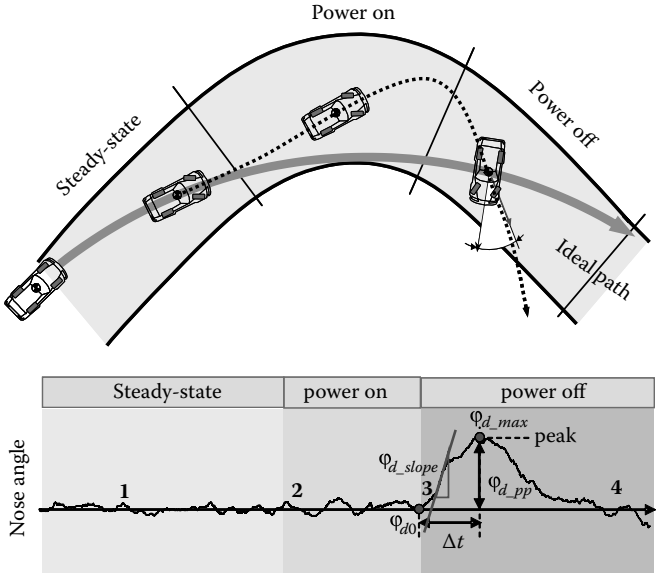


FIGURE 37.69 Nose angle characteristic parameters representing power-off reaction steer.

nose angle change. If φ_{d_0} and φ_{d_max} are the initial and maximum values of φ_d during powering off as shown in Figure 37.69, the peak-to-peak value of φ_d is written as

$$\varphi_{d_pp} = sign(\varphi_{d_max})(\varphi_{d_max} - \varphi_{d_0}) \tag{37.56}$$

As an incremental slope of φ_d, φ_{d_slope} is determined as

$$\varphi_{d_slope} = sign(\varphi_{d_max})\frac{\Delta\varphi_d}{\Delta t} \tag{37.57}$$

where $\Delta\varphi_d/\Delta t$ is the biggest representative slope of φ_d existing between φ_{d_0} and φ_{d_max}.

For the purpose of formulation of the power-off reaction steer, power-off reaction steer coefficient *PS* is defined as

$$PS \equiv w_{ps1}\varphi_{d_slope} + w_{ps2}\frac{\varphi_{d_pp}}{\Delta t_\varphi} + w_{ps3}\varphi_{d_pp} > 0 \tag{37.58}$$

where w_{ps1}, w_{ps2}, and w_{ps3} are the weighing factors of the power-off reaction steer coefficient, and they are also dependent on each driver, vehicle, and test condition. Δt_φ is a time interval between φ_{d_0} and φ_{d_max}.

37.6.2 Analysis

For example, one of the most popular European passenger cars with tires of 215/55R16V was selected to show the process of application of objective formulations representing subjective assessment. With GVW load condition, standard inflated front tire air pressure of 2.4 bar and underinflated rear tire air pressure of 1.4 bar are on purposely used to magnify test results. Through four different handling tests (refer to Figures 37.32, 37.34, 37.37, and 37.22), the subjective assessment and objective measurement were simultaneously carried out.

TABLE 37.7
Subjective Ratings Assessed during the Tests of Lane Changing, Severe Lane Changing, Circuit Driving and Power On/Off in a Turn

Transient Understeer	Transient Oversteer	Power-Off Reaction Steer
6.8	6.5	6.5

For each test, the overall subjective ratings of transient understeer, transient oversteer, and power-off reaction steer were made using a decimal digit rating as shown in Table 37.7. Their bigger values mean better performance with their smaller levels. In addition, δ_{sw}, V_x, V_y, $\dot{\psi}$, a_y, x, and y were measured. Then, κ_d, ω_d, φ_d, and $\dot{\varphi}_d$ of Equations 37.40, 37.38, 37.42, and 37.43 were computed. In order to obtain $\kappa_{d(N)}$, $\omega_{d(N)}$, and $\varphi_{d(N)}$ of Equations 37.50, 37.51, and 37.53, respectively, for simplicity, the objective measurements of quasi steady-state tests were carried out instead of steady-state tests.

In order to quantify the transient understeer, the maximum values of $\Delta\kappa_d$ and $\Delta\omega_d$ of Equations 37.50 and 37.51 are selected for each turn. In Table 37.8, the calculation results of $\kappa_{d(N)}$, κ_d, $\Delta\kappa_d$, $\omega_{d(N)}$, ω_d, and $\Delta\omega_d$ are given. For the purpose of objectification of the transient oversteer, Equations 37.53 and 37.54 are computed using the maximum values of φ_d, $\varphi_{d(N)}$, and $\dot{\varphi}_d$ regardless of steering handwheel angle. Table 37.9 shows the results of $\varphi_{d(N)}$, φ_d, and $\dot{\varphi}_d$ for $\Delta\varphi_d$ and $\Delta\dot{\varphi}_d$ for each peak. For power-off reaction steer, the computational results of φ_{d_0}, φ_{d_max}, φ_{d_slope}, $\varphi_{d_pp}/\Delta t_\varphi$, and φ_{d_pp} are summarized in Table 37.10.

TABLE 37.8
Transient Understeer Parameters

Tests		$\kappa_{d(N)}$ (1/mm)	κ_d (1/mm)	$\Delta\kappa_d$ (1/mm)	$\omega_{d(N)}$ (deg/s)	ω_d (deg/s)	$\Delta\omega_d$ (deg/s)
Lane changing	First turn	5.50	3.00	2.50	11.94	6.81	5.13
	Second turn	−3.15	2.50	5.65	−6.56	5.57	12.13
Severe lane changing	First turn	−17.55	−10.60	6.95	−22.17	−13.74	8.43
	Second turn	10.92	−2.80	13.72	13.21	−3.57	16.78
	Third turn	−32.62	10.30	42.92	−36.11	11.84	47.95
Circuit driving		41.12	25.50	15.62	40.13	26.30	13.83
Powering on/off in a turn		20.67	19.30	1.37	22.16	20.62	1.54

TABLE 37.9
Transient Oversteer Parameters

Tests		$\varphi_{d(N)}$ (deg)	φ_d (deg)	$\Delta\varphi_d$ (deg)	$\dot{\varphi}_d$ (deg/s)	$\Delta\dot{\varphi}_d$ (deg/s)
Lane changing	First peak	1.16	2.22	1.06	18.17	18.17
Severe lane changing	First peak	−1.14	−10.13	8.99	−40.10	40.10
	Second peak	0.55	2.61	2.06	32.74	32.74
Circuit driving		0.36	3.96	3.60	2.77	2.77
Powering on/off in a turn		1.99	2.34	0.34	2.07	2.07

TABLE 37.10
Power-Off Reaction Steer Parameters

Test	t_0(s)	t_{max}(s)	φ_{d_0} (deg)	φ_{d_max} (deg)	Δt_φ(s)	φ_{d_slope} (deg/s)	$\frac{\varphi_{d_pp}}{\Delta t_\varphi}$ (deg/s)	φ_{d_pp} (deg)
Powering on/ off in a turn	1.84	2.23	1.87	2.34	0.39	1.84	1.20	0.47

A further correlation study between subjective assessment and objective measurement is referred to the reference [28]. It is recognized that still there exist some more issues to be resolved for the correlation.

37.6.2.1 Quasi Steady-State

For the purpose of replacing steady-state cornering tests by quasi steady-state tests, weave tests with very slow sweep rate were performed at six different vehicle speeds. For these tests, other driving conditions were kept the same as the main four tests. As shown in Figure 37.70, the measurements of turning curvature $\kappa_{d(N)}$, turning velocity $\omega_{d(N)}$, nose angle $\varphi_{d(N)}$, and nose angle time rate $\dot{\varphi}_{d(N)}$ were obtained as functions of steering handwheel angle δ_{sw} for each vehicle speed V_x. Since the width of the utilized proving ground is limited, the higher vehicle speed was obtained with the less steering handwheel angle.

Figure 37.70 illustrates that the turning curvature and nose angle are dependent on each vehicle speed. However, the turning velocity almost does not depend on the vehicle speed. Within about 90° of steering handwheel angle, they are about linear for 40 kph. For the higher vehicle speed, the narrower linear range of the steering handwheel angle can be achieved.

As shown in Figure 37.70, the nose angle time rate does not have any significant values for all vehicle speeds, instead almost zero. By curve fitting, $\kappa_{d(N)}$, $\omega_{d(N)}$, $\varphi_{d(N)}$, and $\dot{\varphi}_{d(N)}$ are determined as functions of δ_{sw} and V_x.

37.6.2.2 Lane Changing

Lane-changing maneuver contains two consecutive turns as explained in Figure 37.32. For that purpose, one cycle of steering handwheel angle δ_{sw} is applied such as half cycle each in opposite direction as shown in Figure 37.71. Afterward, some more steering correction is sometimes needed to compensate excessive transient oversteer behavior. By applying δ_{sw} in one direction, the first turn from the initial straight line is generated. After the vehicle is laterally displaced, it is necessary to reverse δ_{sw} to keep vehicle path straight in the second lane. By this reverse of δ_{sw}, the second turn is achieved.

For this lane-changing test, δ_{sw}, V_{xd}, V_{yd}, $\dot{\psi}$, and a_{yd} were obtained. Then, κ_d, ω_d, φ_d, and $\dot{\varphi}_d$ are computed using Equations 37.40, 37.38, 37.42, and 37.43. In addition, $\kappa_{d(N)}$, $\omega_{d(N)}$, and $\varphi_{d(N)}$ of Figures 37.72 and 37.73 are computed from the fitting equations of Figure 37.70 as functions of δ_{sw} of Figure 37.71 determined. As results, $\kappa_{d(N)}$, κ_d, $\omega_{d(N)}$, ω_d, $\varphi_{d(N)}$, φ_d, and $\dot{\varphi}_d$ have the same value of δ_{sw} at any instant shown in Figures 37.72 and 37.73.

In Figure 37.72, κ_d and ω_d are shown smaller in the absolute sense than $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively, in the first and second turns. Then, $\Delta\kappa_d$ and $\Delta\omega_d$ of Equations 37.50 and 37.51 become positive in both turns. Therefore, it is confirmed that there exist transient understeer behaviors in the first and second turns during lane changing as shown in Figure 37.32. The understeer tendency in the second turn is shown bigger than that in the first turn.

Figure 37.73 illustrates that each peak of φ_d or $\dot{\varphi}_d$ has occurred due to each reverse steering operation (refer to Figure 37.71). In general, the first peak of φ_d is relatively much smaller than the second peak. It is also located in the zone of the transient understeer (refer to Figure 37.72). In this zone, the transient understeer behavior is believed more dominant than transient oversteer behavior

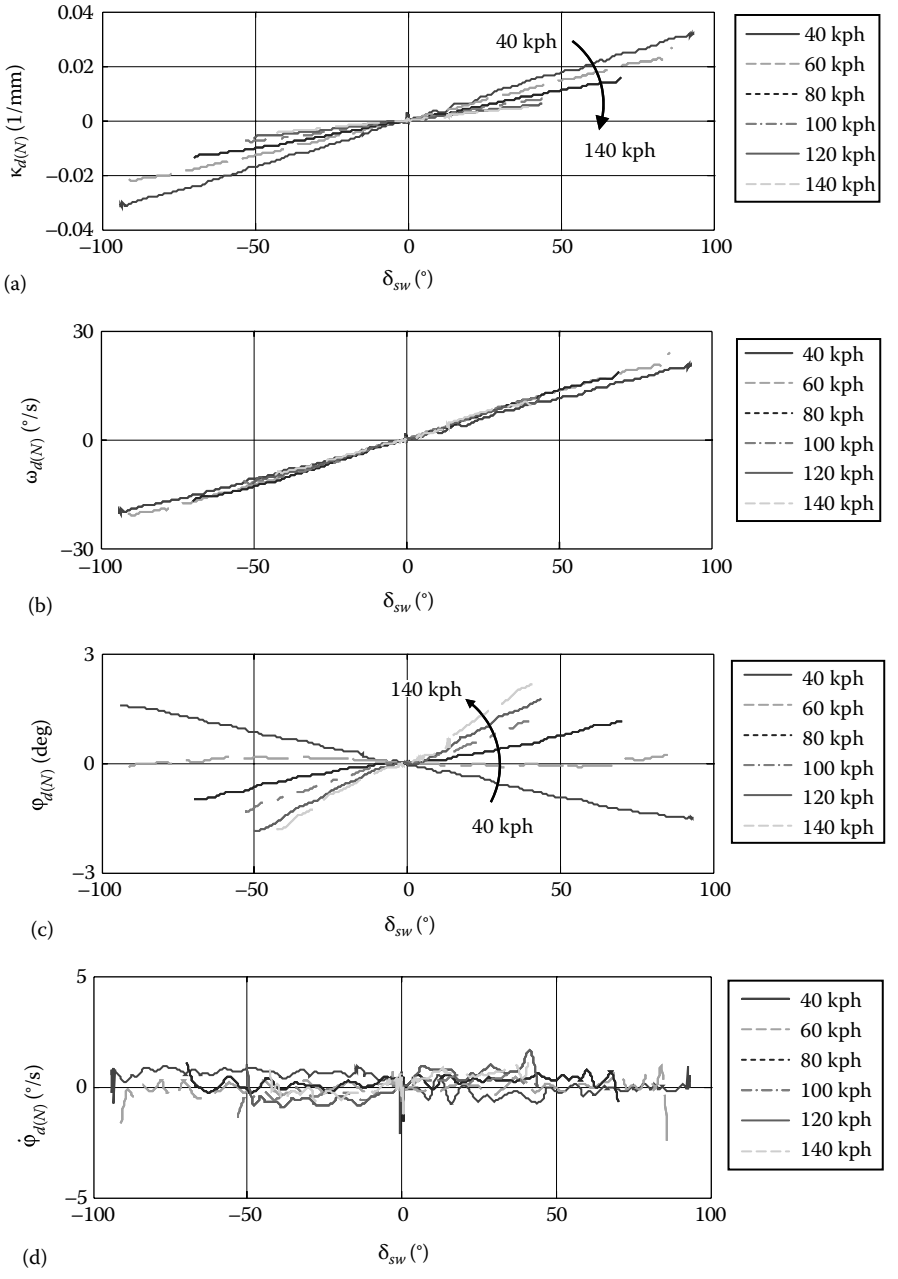


FIGURE 37.70 Quasi steady-state cornering data representing steady-state neutral steer with various vehicle speeds. (a) Turning curvature, (b) turning velocity, (c) nose angle, and (d) nose angle time rate.

so that the oversteer is not subjectively assessed. The second peak explains that the transient oversteer is assessed at the end of lane changing.

After applying the second reverse steer (refer to Figure 37.71), κ_d and ω_d are shown smaller in the absolute sense than $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively, in Figure 37.72. Even though that, the understeer is not subjectively assessed in this region (refer to Figure 37.32). As shown in Figures 37.32 and 37.73, in this region, the oversteer is believed much significant so that only the oversteer is subjectively assessed.

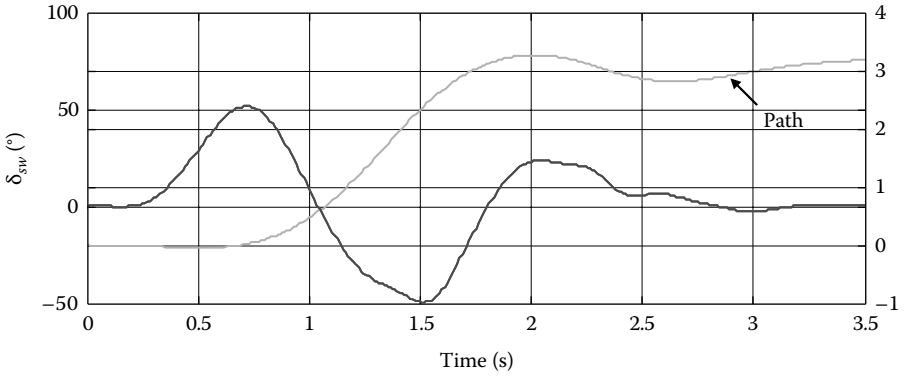


FIGURE 37.71 Steering handwheel angle and vehicle path measured during lane changing.

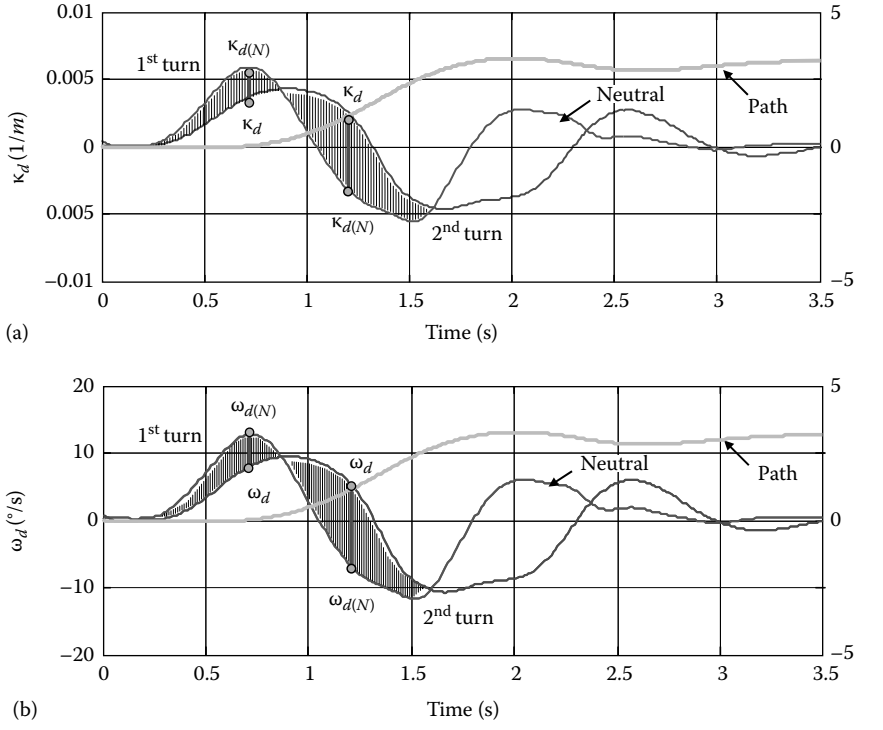


FIGURE 37.72 (a) Turning curvature and (b) turning velocity during lane changing.

37.6.2.3 Severe Lane Changing

In order to accomplish severe lane changing with three consecutive turns, one and half cycles of steering handwheel angle δ_{sw} are first applied (refer to Figure 37.34). Due to unwanted oversteer behavior, afterward, additional steering corrections are needed to correct vehicle path straight as shown in Figure 37.74. Similar to the lane changing, by applying δ_{sw} in one direction, a turn from the initial lane is made.

After a vehicle is laterally displaced closed to the second lane, it is necessary to reverse δ_{sw} and hold that for some moment in order that a vehicle moves back to the original lane again. From this reverse of δ_{sw}, the second turn is achieved. In order to return vehicle path straight in the original lane, afterward, the third turn is achieved by another reverse of δ_{sw} and some more corrections if needed.

Even for the severe lane changing, using the same way of the lane changing, $\kappa_{d(N)}$, κ_d, $\omega_{d(N)}$, ω_d, $\varphi_{d(N)}$, φ_d, and $\dot{\varphi}_d$ are computed as shown in Figures 37.75 and 37.76. Similar to the lane changing,

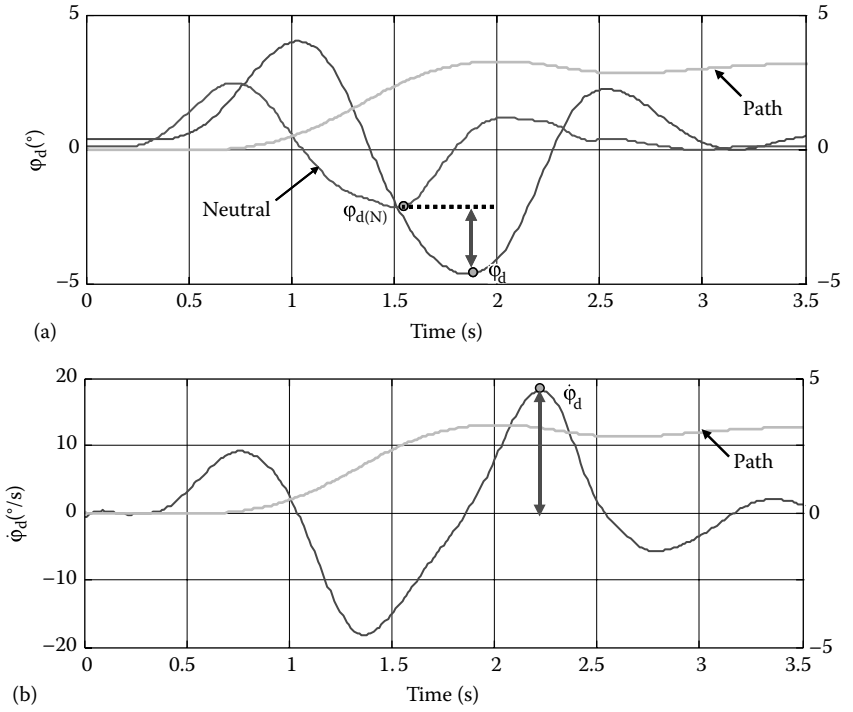


FIGURE 37.73 (a) Nose angle and (b) nose angle time rate during lane changing.

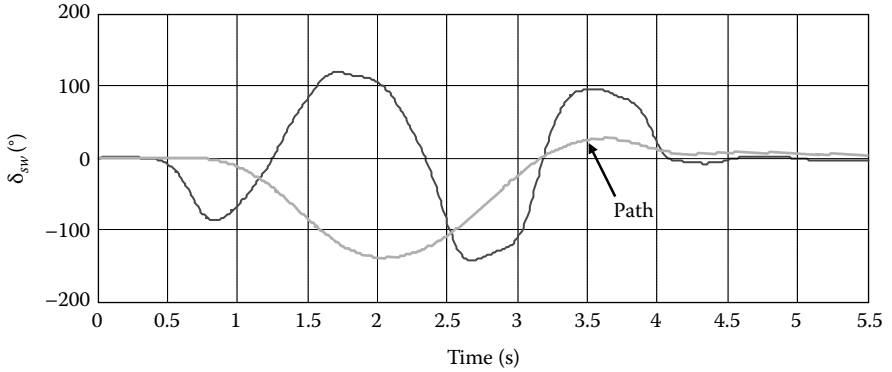


FIGURE 37.74 Steering handwheel angle and vehicle path measured during severe lane changing.

in the absolute sense, κ_d and ω_d are shown smaller than $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively, in each turn in Figure 37.75. Then, $\Delta\kappa_d$ and $\Delta\omega_d$ of Equations 37.50 and 37.51 become positive in all three turns. Therefore, it is also confirmed that there exist transient understeer behaviors in the three turns as shown in Figure 37.34. The understeer tendency in the third turn is the biggest among three turns.

In Figure 37.76, each peak of φ_d, $\varphi_{d(N)}$, or $\dot{\varphi}_d$ is observed after the application of each reverse steer (refer to Figure 37.74). Since the first peak is in general small enough to be ignored, it is believed not meaningful. Furthermore, Figure 37.76 demonstrates that there exist two significant peaks representing transient oversteer behaviors while severe lane changing (refer to Figure 37.34). After the third reverse steer, κ_d and ω_d are shown smaller in the absolute sense than $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively, in Figure 37.75. In this zone, the understeer is believed not perceptually interesting, since the oversteer is more important to be assessed.

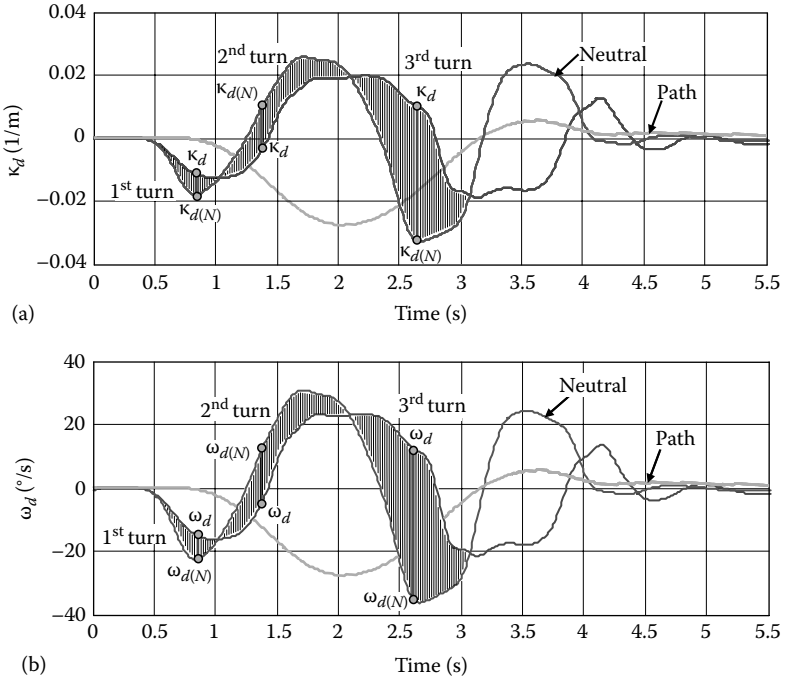


FIGURE 37.75 (a) Turning curvature and (b) turning velocity during severe lane changing.

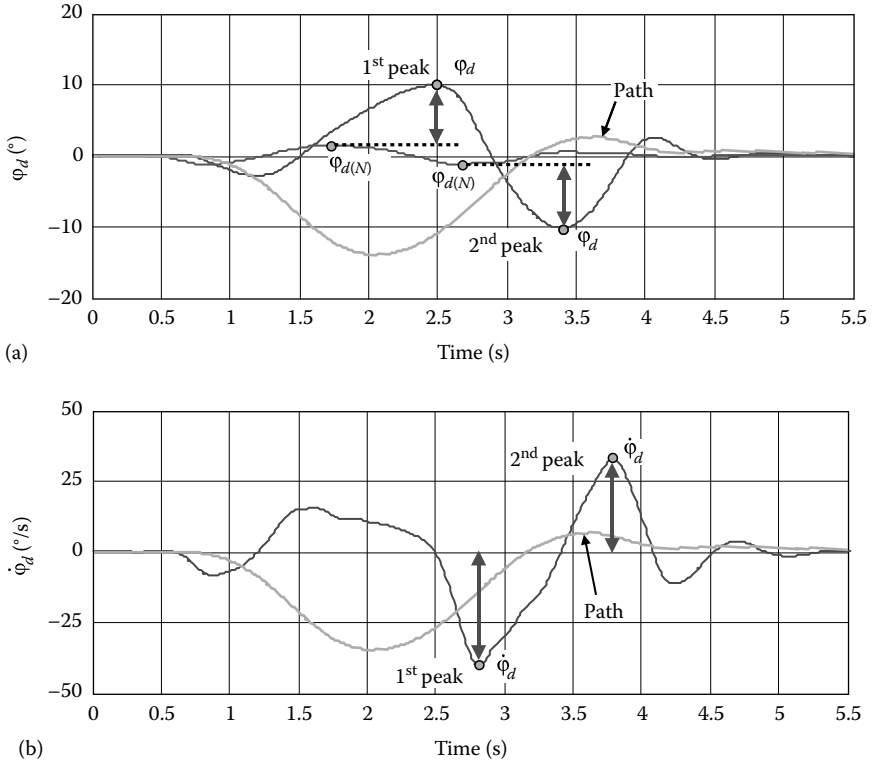


FIGURE 37.76 (a) Nose angle and (b) nose angle time rate during severe lane changing.

37.6.2.4 Circuit Driving

In general, a handling circuit contains many corners with various turning curvatures. By a driver's comment, one corner with the biggest severity was selected for this analysis. Figure 37.77 shows the objective measurement of δ_{sw} at this corner.

For circuit driving, Figures 37.78 and 37.79 are obtained using the same way of the lane changing. Since in Figure 37.78 positive values of $\Delta\kappa_d$ and $\Delta\omega_d$ are observed, transient understeer behavior is expected in this corner. In Figure 37.79, one significant peak of φ_d explains that there also exists transient oversteer tendency (refer to Figure 37.21).

37.6.2.5 Powering On/Off in a Turn

Figure 37.80 shows the objective measurements of steering handwheel angle and vehicle speed during powering on/off in a turn (refer to Figure 37.22). Steering handwheel angle δ_{sw} was maintained

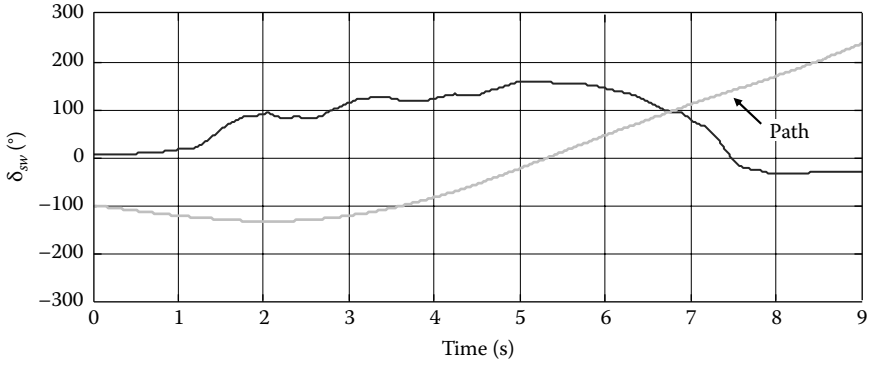


FIGURE 37.77 Steering handwheel angle and vehicle path measured during circuit driving.

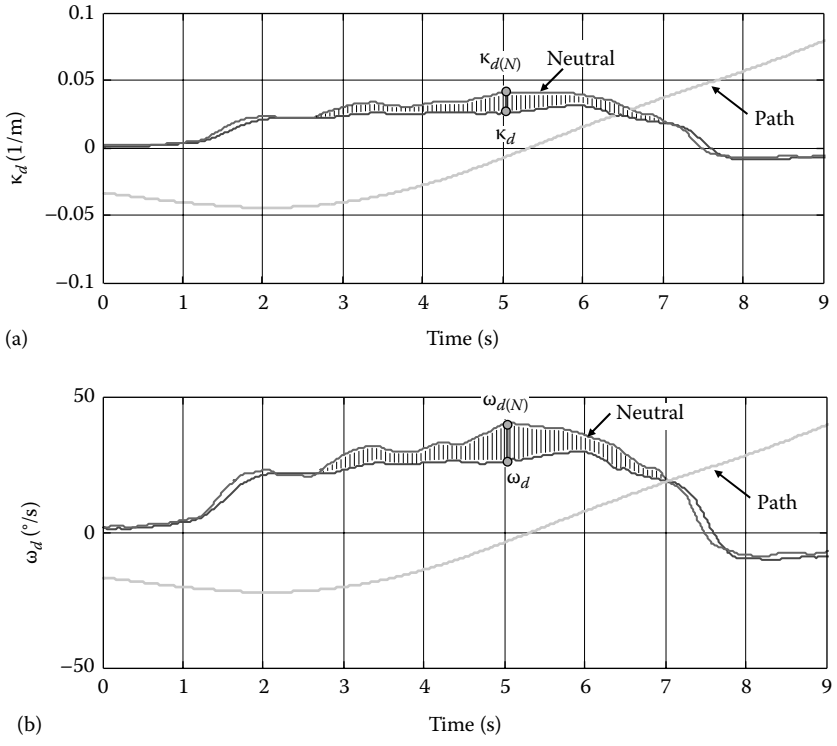


FIGURE 37.78 (a) Turning curvature and (b) turning velocity during circuit driving.

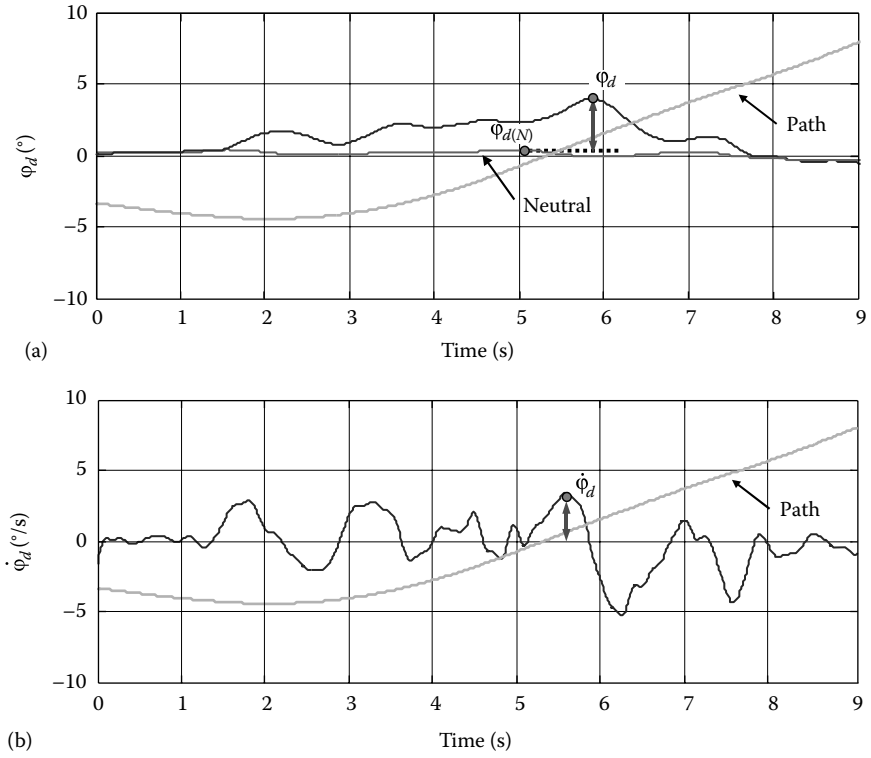


FIGURE 37.79 (a) Nose angle and (b) nose angle time rate during circuit driving.

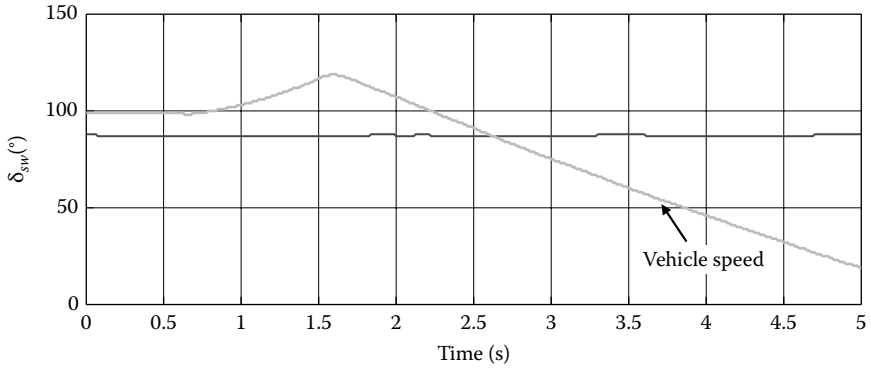


FIGURE 37.80 Steering handwheel angle and vehicle speed measured during powering on/off in a turn.

almost constant during the test. Using the same way of the lane changing, Figures 37.81 and 37.82 are determined for a powering on/off in a turn test.

Even though δ_{sw} is kept constant during the test, $\kappa_{d(N)}$ becomes constant only while steady-state cornering. It is due to vehicle speed change (refer to Figure 37.70). However, both κ_d and ω_d are decreased due to power on as shown in Figure 37.81. This means that a vehicle travels out from the steady-state path. This fact illustrates transient understeer behavior during powering on. Due to power off, in contrast to power on, κ_d and ω_d do not have any meaningful changes relative to $\kappa_{d(N)}$ and $\omega_{d(N)}$, respectively.

Due to constant δ_{sw}, $\varphi_{d(N)}$ also remains constant only while steady-state cornering as shown in Figure 37.82. By power on, only small meaningless disturbances of φ_d and $\dot{\varphi}_d$ from their steady-state values are shown. However, there are significant instantaneous changes of φ_d and $\dot{\varphi}_d$ due to power off.

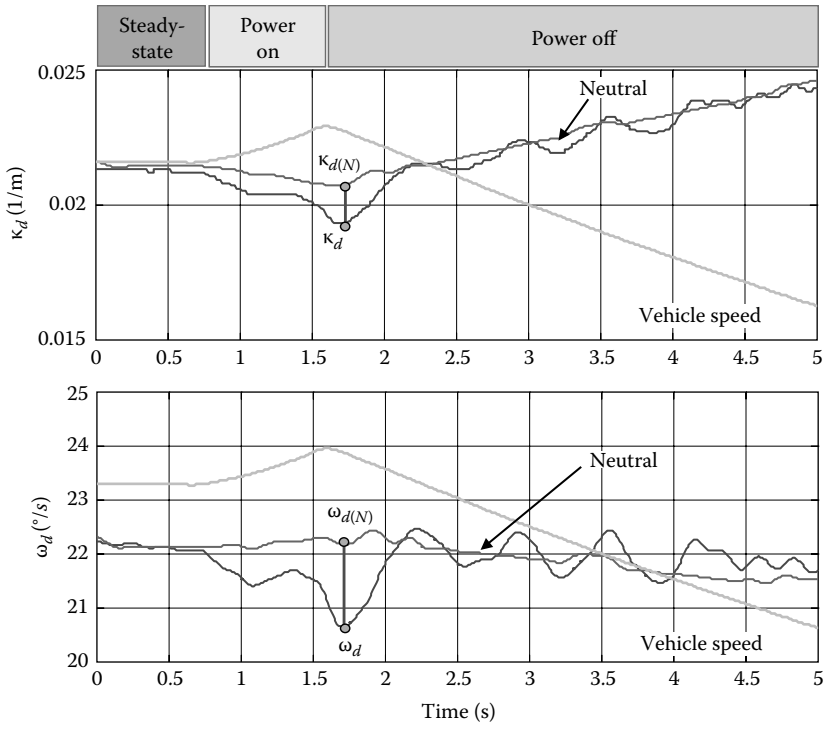


FIGURE 37.81 Turning curvature and turning velocity during powering on/off in a turn.

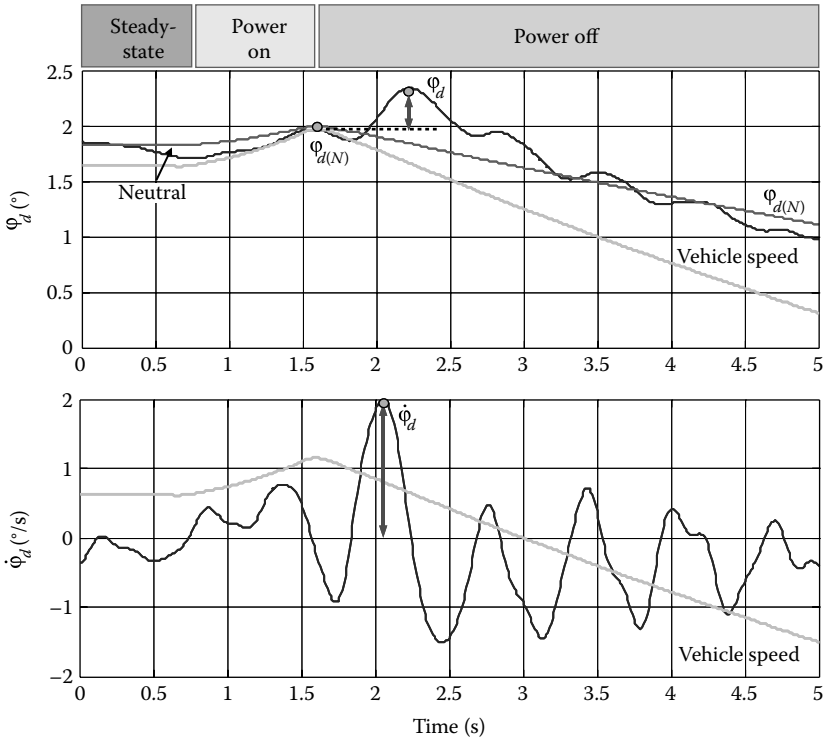


FIGURE 37.82 Nose angle and nose angle time rate during powering on/off in a turn.

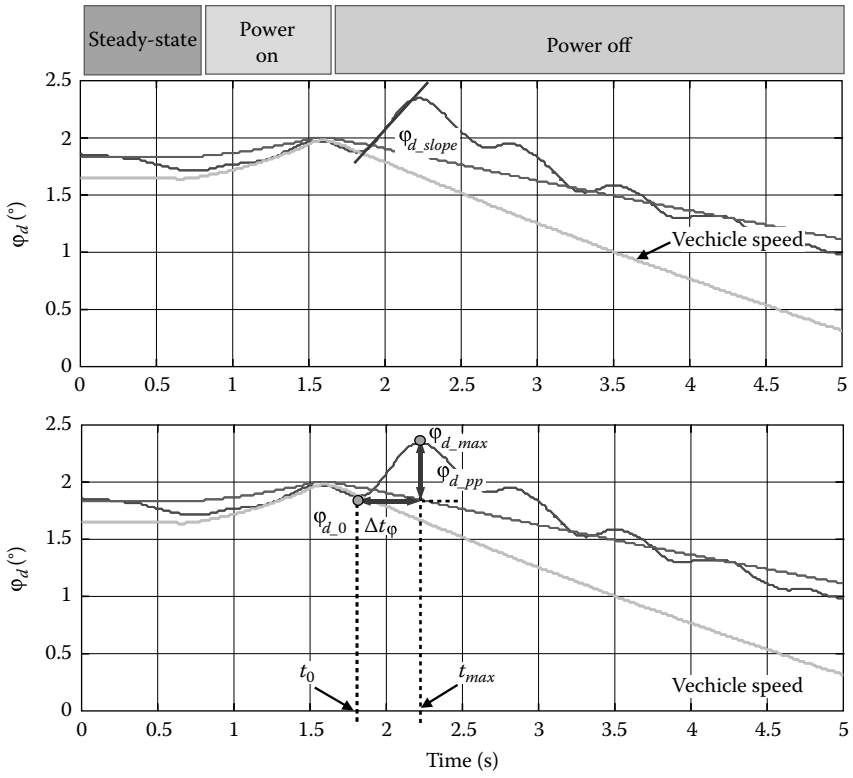


FIGURE 37.83 Characteristic parameters of power-off reaction steer.

These results show that the variation of vehicle nose-in attitude is more influenced by power off. This fact validates transient oversteer behavior during powering off (refer to Figure 37.22).

As the characteristic parameters of power-off reaction steer, φ_{d_slope} and φ_{d_pp} are shown in Figure 37.83. In addition, φ_{d_0}, φ_{d_max}, and Δt_φ are given for the computation of $\varphi_{d_pp}/\Delta t_\varphi$.

37.6.3 Conclusion

Instead of steady-state steer characteristic, nowadays, transient steer characteristic is subjectively assessed as one of handling performance. For that purpose, transient understeer is defined as a perceptual indicator of front axle grip necessary to get controllability in a turn. Transient oversteer and power-off reaction steer are also defined as perceptual measures of rear axle grip. The subjective assessment on the understeer or oversteer is significantly dependent on vehicle path or vehicle attitude, respectively. However, the understeer or oversteer does not much depend on the vehicle attitude and vehicle path, respectively. The power-off reaction steer is subjectively assessed based on a driver's perception on vehicle attitude impulsive change.

The transient understeer is formulated as a function of turning curvature and turning velocity, while the transient oversteer is written as a function of nose angle and nose angle time rate. The power-off reaction steer is expressed as a function of the incremental slope, (effective) durational change, and peak-to-peak value of the nose angle. These formulations are applied through four handling tests. The test results show the understeer, oversteer, and power-off reaction steer quantitatively described in figures.

In order to illustrate subjective tests necessary for each subjective assessment, for example, a cross-functional matrix between characteristics and tests regarding handling performance is given in Table 37.11. In Table 37.12, the summary of each subjective assessment is described. In the table,

TABLE 37.11
Cross-Functional Matrix between Characteristics and Tests Regarding Handling Performance

		Straight Ahead				Steering					Lane Changing		Cornering			
Performance	Maneuver	Free Steering	Fixed Steering	Powering On/Off	Braking	On-Center Steering	Off-Center Steering	Progressive Steering	Limit-Grip Steering	Flick Steering	Lane Changing	Severe Changing	Circuit Driving	Powering On/Off in a Turn	Braking in a Turn	Rollover Cornering
Straightability	Residual pull	✓														
	Running straight		✓													
	Torque steer			✓												
	Braking straight				✓											
Steerability	On-center					✓										
	Steering response						✓									
	Linearity							✓								
	Steering precision												✓			
	Steering angle												✓			
	Steering addition								✓							
	Steering effort						✓						✓			
	Returnability									✓			✓			

| | | |

TABLE 37.12
Summary of Subjective Assessments Regarding Handling Performance

Performance		Description	What Is Evaluated	How to Test
Straightability	Residual pull	Lateral drift characteristic of a vehicle caused due to residual steering torque with free steering at constant speed	Lateral deviation	Free steering in straight ahead
	Running straight	Straight-running ability without any drift, wander, or sloppiness with fixed steering at constant speed	Drift, wander, and sloppiness	Fixed steering in straight ahead
	Torque steer	Lateral drift characteristic of a vehicle due to power on in straight ahead with fixed steering	Lateral deviation	Powering on/off in straight ahead
	Braking straight	Straight-braking ability without any disturbance of vehicle attitude or path when a vehicle instantaneously stops complete	Attitude or path deviation	Braking in straight ahead
Steerability	On-center	Steering torque and vehicle path change in response to a very slow steering input at and around the center of steering	Steering torque	On-center steering
	Steering response	Time-dependent vehicle path change against a transient steering input	Vehicle path change	Off-center steering
	Linearity	Linearity characteristic of trajectory of vehicle path against a gradually incremental steering input	Profile of vehicle path	Progressive steering
	Steering precision	Overall impression on how accurately a vehicle follows an intended path with a series of steering maneuvers	Path following Easiness to control	Circuit driving
	Steering angle	Output-to-input ratio of a steering input in getting the same path	Steering handwheel angle	Circuit driving
	Steering addition	Additional grip produced by an extra steering input at grip limit in a turn	Vehicle path change	Limit-grip steering
	Steering effort	Steering handwheel torque required to change the direction of vehicle path	Steering handwheel torque	Off-center steering, circuit driving
	Returnability	Quickness and precision of steering handwheel movement returning to its neutral position	Oscillation of steering handwheel angle	Flicking steering, circuit driving

TABLE 37.12 (continued)
Summary of Subjective Assessments Regarding Handling Performance

Performance		Description	What Is Evaluated	How to Test
Controllability	Balance FA/RA	Balance FA/RA	Vehicle path relative to that of NS Vehicle attitude relative to that of NS	Lane changing, severe lane changing circuit driving, powering on/off in a turn
	Power-off reaction	Steer characteristic due to an instantaneous load transfer from the rear axle to a front axle while powering off with fixed steering in a turn	Vehicle path relative to that of NS Vehicle attitude relative to that of NS Vehicle attitude impulsive change	Powering on/off in a turn
	Grip limit	Attainable maximum speed in still controlling a vehicle in lane changing or cornering	Speed limit, vehicle path Vehicle attitude	Lane changing, severe lane changing Circuit driving
	Controllability at limit	Overall impression on how easy to control a vehicle at grip limit	Vehicle path Vehicle attitude	Lane changing, severe lane changing Circuit driving, powering on/off in a turn
	Driving power	Traction ability to speed up while powering on in the exit zone of a corner	Vehicle speed increase	Circuit driving, powering on/off in a turn
Stability	Yaw stability	Ability to keep down the yaw oscillation of a vehicle	Oscillation of vehicle path	Flick steering, lane changing, severe lane changing
	Rear axle stability	Ability to restrict the sideslip oscillation of a vehicle	Oscillation of vehicle attitude	Flick steering, lane changing, severe lane changing, powering on/off in a turn
	Roll stability	Ability to depress the roll oscillation of a vehicle	Oscillation of vehicle body motion	Flick steering, lane changing, severe lane changing
	Braking stability	Ability to stop a vehicle without any disturbance of vehicle attitude or path while full braking in a turn	Disturbance of vehicle attitude or path	Braking in a turn
	Rollover stability	Ability to restrain a vehicle from its rollover motion.	Debeading Tire lift-off	Rollover cornering
	Bump steer	Steer instability caused due to the jounce and rebound of each front suspension	Vehicle path change	Circuit driving

TABLE 37.13
Conditions of Subjective Tests Regarding Handling Performance

	Test	Speed (kph)	Steering (deg)	Powering	Braking
Straight ahead	Free steering	100 (high gear)	Free	Constant	NA
	Fixed steering	Up to limit	Fixed	Constant	NA
	Powering on/off	40 (low gear) 80 (high gear)	Fixed	Full on and off	NA
Steering	Braking	Normal driving	Hands-on	NA	Full
	On-center steering	Up to limit	Miniature slow perceptual ramp	Constant	NA
	Off-center steering	Up to limit	$\pm(\delta_{sw0} + \Delta\delta_{sw})$ Perceptual step	Constant	NA
	Progressive steering	Up to limit	High and mid perceptual ramp	Constant	NA
	Limit-grip steering	Up to limit	$\pm(\delta_{sw0} + \Delta\delta_{sw})$ Perceptual ramp	Constant	NA
	Flick steering	Up to limit	±90 Perceptual impulse	Constant	NA
Lane changing	Lane changing	Up to limit	±45 Perceptual sinusoidal	Constant	NA
	Severe lane changing	80	±90 Perceptual sinusoidal	Off	NA
Cornering	Circuit driving	Up to limit	Transient	Up to full	Up to full
	Powering on/off in a turn	Up to limit	Fixed	Full on and off	NA
	Braking in a turn	Up to limit	Fixed	NA	Full
	Rollover cornering	Dependent on each test procedure			NA

what is evaluated and how to test are also briefly explained. Table 37.13 outlines the detailed conditions of speed, steering, powering on/off, and braking for each subjective test.

Figure 37.84 shows a hierarchical diagram from vehicle motion variables to performance through subjective assessment variables. Some of the vehicle motion variables are measured during numerous subjective tests. The objective measurement variables yield the physical perception variables defined with respect to a driver's point of view. The physical perception variables describe the subjective assessment variables in mathematical formulations.

The subjective assessment variables explain the driver's perceptual judgment on characteristics. The characteristics are the subgroups of capabilities. Finally, the capabilities are represented by performance. For an example, Table 37.14 illustrates the objective measurement variables, physical perception variables, and subjective assessment variables regarding transient understeer, transient oversteer, and power-off reaction steer.

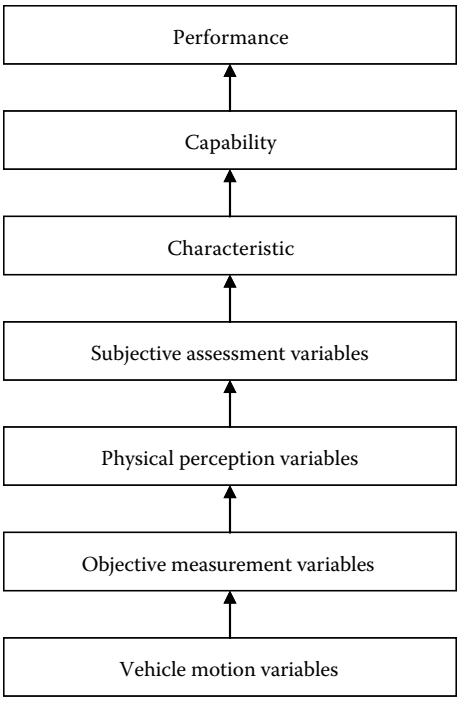


FIGURE 37.84 Hierarchical diagram from vehicle motion variables to performance through subjective assessment variables.

TABLE 37.14
Variables of Objective Measurement, Physical Perception, and Subjective Assessment Regarding Understeer, Oversteer, and Power-Off Reaction Steer

Performance	**Handling**		
Capability	**Controllability**		
Characteristics	**Transient Understeer**	**Transient Oversteer**	**Power-Off Reaction Steer**
Subjective assessment variables	Vehicle speed Steering handwheel angle Vehicle path	Vehicle speed Steering handwheel angle Vehicle attitude	Vehicle speed Vehicle attitude impulsive change
Physical perception variables	Vehicle speed Steering handwheel angle Turning curvature Turning velocity	Vehicle speed Steering handwheel angle Nose angle Nose angle time rate	Vehicle speed Nose angle incremental slope Nose angle durational change Nose angle peak-to-peak value
Objective measurement variables		Longitudinal velocity Lateral velocity Steering handwheel angle Yaw rate Lateral acceleration	

ACKNOWLEDGMENT

Special thank is owed to Dr. Hwi Joong Kim, the previous CTO of Hankook Tire, for his encouragement in the beginning and continuous support. I would like to thank Mr. Roger G. Becker, Lotus, for his open discussions; John P. Whitehead for a specially designed training program with MIRA; and to Mr. Franz-Jurgen Dressen, Ford, for his clear demonstration and discussion. Particular thanks are also due to Jung-Sik Kim for his successive help until the completion of this work.

REFERENCES

1. Segel, L., *Theoretical Prediction and Experimental Substantiation of the Response of the Automobile to Steering Control*, Automobile Division, the Institution of Mechanical Engineers, London, 1956.
2. Whitcomb, D.W. and Milliken, W.F., *Design Implications of a General Theory of Automobile Stability and Control*, Automobile Division, the Institution of Mechanical Engineers, London, 1956.
3. Pacejka, H.B., Simplified analysis of steady-state turning behaviour of motor vehicles. Part 1. Handling diagrams of simple systems, *Vehicle System Dynamics* 2, 161–172, 1973.
4. Pacejka, H.B., Simplified analysis of steady-state turning behaviour of motor vehicles. Part 2. Stability of the steady-state turn, *Vehicle System Dynamics* 2, 173–183, 1973.
5. Pacejka, H.B., Simplified analysis of steady-state turning behaviour of motor vehicles. Part 3. More elaborate systems, *Vehicle System Dynamics* 3, 185–204, 1973.
6. Wong, J.Y., *Theory of Ground Vehicles*, 3rd edn., John Wiley & Sons, Inc., New York, 2001.
7. Milliken, W.F. and Milliken, D.L., *Race Car Vehicle Dynamics*, SAE International, Pennsylvania, Inc., 1995.
8. Gillespie, T.D., *Fundamentals of Vehicle Dynamics*, SAE International, Pennsylvania, Inc., 1992.
9. Dixon, J.C., *Tires, Suspension and Handling*, SAE International, Pennsylvania, Inc., 1996.
10. Griffin, M.J., *Handbook of Human Vibration*, Academic Press, London, U.K., 1990.
11. ISO, Road vehicles—Vehicle dynamics and road-holding ability—Vocabulary, International Standard, ISO 8855:1991(E/F), 1991.
12. ISO, Passenger cars—Steady-state circular driving behavior—Open-loop test procedure, International Standard, ISO 4138:1996(E), 1996.
13. ISO, Road vehicles—Lateral transient response test methods—Open-loop test methods, International Standard, ISO 7401:2003(E), 2003.
14. ISO, Road vehicles—Transient open-loop response test method with one period of sinusoidal input, International Standard, ISO/TR 8725:1998(E), 1998.
15. ISO, Road vehicles—Transient open-loop response test method with pseudo-random steering input, International Standard, ISO/TR 8726, 1988.
16. ISO, Passenger cars—Braking in a turn—Open-loop test procedure, International Standard, ISO 7975:1996(E), 1996.
17. ISO, Passenger cars—Test track for a severe lane-change manoeuvre—Part 1: Double lane-change, International Standard, ISO 3888–1, 1999.
18. ISO, Mechanical vibration and shock: Evaluation of human exposure to whole-body vibration, International Standard, ISO 2631-1:1997, 1997.
19. Kruger, H.P., Neukum, A., and Schuller, J., A workload approach to the evaluation of vehicle handling characteristics, Society of Automotive Engineers, No. 2000-01-0170, 2000.
20. Stephens, P. and Kohn, H.J., Influence of the tyre on subjective handling at the limit, In *Vehicle Performance—Understanding Human Monitoring and Assessment*, pp. 177–195, 1999.
21. Godthelp, A.P., de Vos, J., and Käppler, W.D., Subjective and objective assessment of manual, supported, and automated vehicle control, In *Vehicle Performance—Understanding Human Monitoring and Assessment*, pp. 97–120, 1999.
22. Crolla, D.A. and Chen, D.C., Vehicle handling assessment using a combined subjective-objective approach, Society of Automotive Engineers, No. 980226, 1998.
23. Norman, K., Objective evaluation of on-center handling performance, Society of Automotive Engineers, No. 840069, 1984.
24. Farrer, D.G., An objective measurement techniques for the quantification of on-centre handling quality, Society of Automotive Engineers, No. 930827, 1993.
25. Salaani, M.K., Experimental steering feel performance measures, Society of Automotive Engineers, No. 2004-01-1074, 2004.

26. Walter, S.L., Quantitative measurement of handling characteristics of tires and/or vehicle/tire combinations, U. S. Patent No. 4969212, 1990.
27. Gauthier, P.N. and Ludlow, W., System and method for testing deflated tire handling, U. S. Patent No. 6580980B1, 2003.
28. Gim, G. and Kim, J., Objectification of subjective assessment on transient steer characteristic, FISITA, No. F2004F071, 2004.
29. Wood, J.J. and Leatherwood, J.D., A new ride quality meter, Society of Automotive Engineers, No. 850981, 1985.
30. Park, W.H. and Wambold, J.C., Objective ride quality measurement, Society of Automotive Engineers, No. 760360, 1976.
31. Mehta, N.C., Subjective and objective ride evaluations of commercial vehicles, Society of Automotive Engineers, No. 810046, 1981.
32. Norsworthy, T.H., The correlation of objective ride measures to subjective jury evaluations of class 8 COE vehicles, Society of Automotive Engineers, No. 850985, 1985.
33. Kudritzki, D.K., Analysis of ride comfort considering driver assessment, In *Vehicle Performance*, J.P. Pauwelussen (ed.), pp. 196–217, 1999.
34. Hassan, R. and McManus, K., Heavy vehicle ride and driver comfort, Society of Automotive Engineers, No. 2001-01-0386, 2001.
35. Strong, J.R., Statistical methods for evaluating truck ride quality measures, Society of Automotive Engineers, No. 850986, 1985.
36. Ushijima, T. and Kumakawa, S., Objective harshness evaluation, Society of Automotive Engineers, No. 951374, 1995.
37. Lee, S.K. and White, P.R., Application of wavelet analysis to the impact harshness of a vehicle, *Proceedings of the Institution of Mechanical Engineers* 214(11), 1331–1338, 2000.
38. Kennedy, D.C., Patten, W.N., and Ratliff, R.T., *Impact Harshness of Small Interstate Bumps: Objective and Subjective Testing*, DSC-Vol. 52, Advanced Automotive Technologies, ASME, New York, 1993.
39. Amman, S., Gu, P., Mouch, T., Meier, R., and Greenburg, J., Sound and vibration contribution to the perception of impact harshness, Society of Automotive Engineers, No. 2005-01-1499, 2005.
40. Gim, G. and Kim, J., Vehicle dynamics of MIRA, Hankook tire, Technical Report No. VD-R98-2, 1998.
41. Gim, G. and Kim, J., Subjective assessment of TÜV for dry handling/stability, Hankook tire, Technical Report No. VD-R98-3, 1998.
42. Gim, G., Subjective assessment of Ford SVC for dry handling/stability, Hankook tire, Technical Report No. VD-R98-6, 1998.
43. Gim, G., Kim, J., Oh, S., and Kim, J., Modern vehicle dynamics: Subjective test of handling, Hankook tire, Technical Report No. VD-R01-01, 2004.
44. SAE, *Subjective Rating Scale for Vehicle Handling*, *J1441*, Society of Automotive Engineers, Pennsylvania, 1998.
45. Smith, C., *Tune to Win: The Art and Science of Race Car Development and Tuning*, Aero Publishers, Inc., Fallbrook, CA, 1978.
46. Bentley, R., *Speed Secrets: Professional Race Driving Techniques*, MBI Publishing Company, Osceola, WI, 1998.
47. Lopez, C., *Going Faster: Mastering the Art of Race Driving*, Robert Bentley, Inc., Cambridge, MA, 1997.
48. Forkenbrock, G.J., O'Harra, B.C., and Elsasser, D., A demonstration of the dynamic tests developed for NHTSA's NCAP rollover rating system: Phase VIII of NHTSA's light vehicle rollover research program, NHTSA DOT HS 809 705, 2004.
49. Jang, H.K. and Griffin, M.J., Effect of phase, frequency, magnitude and posture on discomfort associated with differential vertical vibration at the seat and feet, *Journal of Sound and Vibration* 229(2), 273–286, 2000.
50. Fey, B., *DATA POWER: Using Racecar Data Acquisition*, Towery Publishing, Inc., Memphis, TN, 1993.
51. SAE, *Vehicle Dynamics Terminology*, *SAE J670e*, Society of Automotive Engineers, Warrendale, Publishing, 1976.

38 Driver Models in Automobile Dynamics Application

Manfred Ploechl and Johannes Edelmann

CONTENTS

38.1 Introduction 1477
38.2 Applications with Focus on the Vehicle 1479
38.2.1 General Remarks 1479
38.2.2 Virtual Test Driver Models 1480
38.3 Applications with Focus on the Driver 1495
38.3.1 Understanding Driver and (Individual) Driving Behavior 1495
38.3.2 Path and Speed Planning, Optimized Driver/Driving Behavior 1500
38.4 Applications with Focus on Vehicle/Driver Combination 1503
38.4.1 Vehicle Handling Dynamics with Respect to the Driver 1504
38.4.2 Accident Avoidance, Active Safety, Driver Support Systems 1506
38.5 Applications with Focus on the Environment/Traffic 1507
38.6 Conclusion 1512
References 1512

Understanding the driver of an automobile has been attractive to researches from many different disciplines for more than half a century. On the basis of their acquirements, models of the (human) driver have been developed to better understand, analyze, and improve the combined couple of driver and automobile.

Due to distinctive demands on the models in accordance with different kinds of applications, a variety of various driver models are available. An overview of driver models is given with respect to their application and different methodical modeling approaches. The emphasis is put on the interest of engineers, who generally focus on the automobile (like design and optimization of vehicle components and the overall vehicle dynamics behavior) by applying their approved (mathematical) methods. Nonetheless, a brief look beyond is added to better complete the view on the involved task of driving and driver modeling for automobile dynamics application.

38.1 INTRODUCTION

In the early 40s of the last century, a systematic investigation of the dynamics of automobiles has been initiated [1]. The real automobile was mapped into a mathematical model, represented by differential equations, and a basic understanding of the vehicle system dynamics (VSD) was attained. It was not before the mid-1950s that the driver of the automobile was brought into view [2], in the 1960s e.g., [3–7]. Although it was already stated in the 1960s that automobile and driver form a closely coupled man–machine system and must consequently be treated as a whole [8], the main concern and focus—with the enormous amount of scientific publications in mind—remained on the vehicle ("without the driver"). And considering still the success of the development of automobiles, one may even see no deficit. With the introduction of systems controlling the dynamics and stability of the automobile, driver assistance

systems, and other systems taking over some control of the driver, all doubts should be resolved on the need of capable driver models, and many publications can be noted in this respect.

Further progress in understanding the human driving behavior was made since then, by putting engineers in touch with other disciplines like physiology and psychology (e.g., [9–11]), and a number of demands on models that aim to mimic the automobile driver have been addressed. Some of the key demands to be mapped in a driver model are listed briefly in the following:

- General human driver skills and characteristics (such as visual, vestibular, tactile, auditory), information reception, perception and processing, neuromuscular dynamics with thresholds, time delays and limitations, preview, prediction/anticipation, adaptation/learning, planning capabilities (path and speed adjustment)
- Specific types of drivers with respect to experience, age, willingness to take and accept risks, etc.
- Portraying features like concentration, tiredness, stress, emotions

Whether a model can be called a driver model of a human and reflects human driving behavior in a sufficient way or not will not be addressed here, but at least some of the aforementioned demands must be reflected in a model to distinguish from "automatic driving controllers", which are not covered here. Despite this restriction, a large number of articles with driver models have been published so far; a very extensive and thorough list of publications up to 1997 with many details on the human driver is given by Jürgensohn in [12]. More recent papers up to 2003 and a valuable insight into the understanding and modeling of the human driver are included in a state-of-the-art article on this topic by MacAdam [13].

Quite a lot of published driver models have basic ideas in common; they often differ in the way the previously listed demands are realized mathematically. But there are also more principal differences that usually result from differently emphasized demands dependent on the application. The applied methodical approach can be used to categorize driver models, for instance, into

- Mathematical methods, using control theory (transfer functions in frequency domain, optimal control, adaptive control, predictive control), identification theory, fuzzy logic control (FLC), neural networks (NNs), stochastic methods, and hybrid approaches, which generally result in (non)linear algebraic or differential equations
- Descriptive methods, resulting in classifications, schemes, descriptions, etc.

Another way of categorizing driver models is to have a look at their applications. Indeed, it is quite obvious, and this contribution has to prove that there is (can be) no unique driver model, but there are only driver models especially tailored to meet the specific demands of each application [14]. Although categorizing driver models by application seems to be reasonable, this intention is not as straightforward as it seems to be. On the one hand, a single driver model may be used for several different applications; on the other hand, applications themselves must be classified and cannot always be clearly separated from each other. To overcome this problem, the authors have chosen comprehensive categories of applications, which are refined at the beginning of each section. Within a category, models are combined to methodic bundles.

Although automobile and driver form a couple, the aim and reason of investigation are often focused either on the vehicle or on the driver, but still there are applications where the couple is not split. After defining driving task and environmental situation, which may include other automobiles with drivers (traffic), the main concern of analysis may be

- On the vehicle: Design and tuning of vehicle components, of the overall vehicle; vehicle dynamics; active safety matters with respect to onboard systems
- On the driver: Understanding driver and (individual) driving behavior in general, path and speed planning features, optimized driver behavior, subjective driving perception

- On the overall (combined) system: Reconstruction of accidents, accident analysis/prevention, learning tasks, evaluation of drivability and handling, driver support systems, active safety
- On the environment/traffic: Influence on/of other drivers in simulator or in traffic, traffic flow, general understanding of the "system" traffic

In this way, also the following sections are arranged. As comprehensive literature surveys are already available [12,13,15,16], often referred to papers are selected here to give a basic insight into "typical" driver models, considering the variety of different applications and mathematical approaches.

38.2 APPLICATIONS WITH FOCUS ON THE VEHICLE

38.2.1 General Remarks

A lot of vehicle dynamics application is done by engineers and they generally have their focus and interest on the vehicle itself. A model of the driver is mainly required to resolve given driving tasks that are the means for closed-loop tests and simulations. One may call this kind of driver a "virtual test driver" [17], as the main demand for this kind of applications on the driver model is to be able to follow a prescribed route at a given or self-imposed speed by operating a virtual prototype of the vehicle. Output quantities of the driver model are steering wheel angle/torque, accelerator or brake pedal position, and gear.

Typical interests of investigation are

- Vehicle component design
- Vehicle dynamics analysis, such as (closed-loop) handling behavior
- Overall vehicle stability analysis
- Design of onboard controls that influence vehicle dynamics and stability

For vehicle component design, load spectra are required, in particular to estimate life of fatigue. When measurements are not (yet) available, simulation is a tool to quantify them. Often the driver model is here more like a "driver robot" than a "test driver" when its task is simply kept to tracking at given speed profiles. But one has to keep in mind that the input from a technical driver and a human driver may be quite dissimilar even if this becomes not obvious from the motion of the vehicle as a result of the integrating behavior of the vehicle [17].

The "test driver" model needs enough complexity to represent both learnt patterns, which are considerably important in car driving, and correction elements. Driver models that include some level of anticipation or prediction or an internal model [13] and some level of compensation capability may well fit for this purpose.

Closely related to vehicle component design is simulating and optimizing (closed-loop) handling behavior—a typical task of vehicle dynamics analysis. In particular, effects of specific design parameters, for example, of the suspension, the steering system, or the tires, on handling and stability behavior are analyzed. However, when it comes to (subjective) evaluation of driveability and handling, the human test driver can still hardly be replaced (see Section 38.4 [12]).

Stability of the vehicle has long been studied without the driver (in open-loop tests) although the driver can appear to be a destabilizing part of the system. The introduction of control systems, that influence vehicle dynamics additionally to the driver, stimulated closed-loop stability analysis that consequently includes a driver model. For the design of this kind of controllers, the driver must be considered, either in outdoor or in simulator tests and in simulations, to account for a possible interference and coordination of both "controllers".

Certain types of virtual test driver models are used quite collectively for all these kinds of applications, mostly with control (theory) strategies, but also other methodical approaches can be noticed. Individual properties of the driver are often disregarded; nonetheless, models representing classes of particular types of drivers (with respect to experience, age, skills, etc.) are favorable.

Models of the automobile tend to be more complex and better validated than driver models and are nowadays often assembled with multibody system (MBS) or VSD software (ADAMS, CarSim, LMS, SIMPACK, veDYNA, and others). This may be the reason why driver models are offered from some MBS software packages as well, as closed-loop simulation will become quite easy from the user point of view. The basic idea of three of these driver models is reviewed here, and another in [18].

Some typical representatives of their class are presented in the following taking into account different mathematical approaches.

38.2.2 Virtual Test Driver Models

One may consider "driving an automobile" intuitively as a control task, and this may be a reason why control theory–based driver models have been successful since the first driver models have been introduced. Most of these models have been developed to control the lateral vehicle dynamics (for previous reviews, see [12,13,15,19]). Longitudinal dynamics is often simply kept to following a given or optimized speed profile, rather independent from the steering task of the driver, considering more the capabilities of the vehicle than those of the driver.

A look back to 1953, when Kondo [2,20] started with driver modeling in Japan, may be helpful to illustrate some of the roots that many following models up to now are (although not directly) obliged to [21]. Based on the two-wheel vehicle model on a straight line, running at constant speed with side wind disturbances, Kondo's driver steers in a way that a point, a certain distance along the vehicle centerline ahead, is going to coincide with the sight point, which is the preview distance L ahead on the reference course (see Figure 38.1).

From a control theory point of view, the steering strategy can be understood as the attempt of the driver to diminish Δy_p a distance L ahead of the vehicle. In terms of transfer functions, this can be illustrated in a more general form by the control loop of Figure 38.2, where $P(s)$ represents the preview strategy of the driver for the desired path y_d with path input y_0, $H(s)$ the control characteristics of the driver, $G(s)$ the transfer function(s) of the vehicle, and $B(s)$

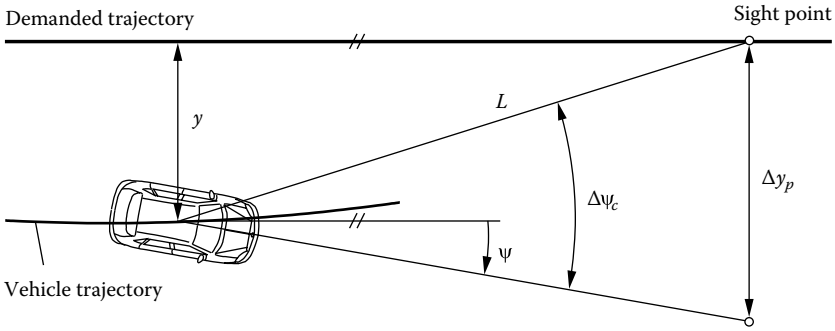


FIGURE 38.1 Kondo's "shaft" driver model.

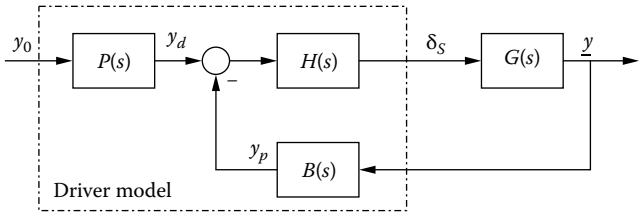


FIGURE 38.2 Basic control loop of preview tracking models.

the feedback function respectively, which is the prediction function of the vehicle motion. According to Figure 38.1, Δy_p can be seen as

$$\Delta y_p(t) \approx y(t) + L\psi(t) = y(t) + T_p v\psi(t) \approx y(t + T_p) \tag{38.1}$$

with the preview time $T_p = L/v$ following Kondo's suggestion in [20], vehicle speed v, yaw angle $\psi(t)$, and deviation $y(t)$. Consequently, Δy_p can indeed be interpreted as predicted lateral deviation $y(t + T_p)$ determining the prediction function $B(s) = e^{T_p s}$ in Figure 38.2 with $y_p = \Delta y_p$. With $P(s) = e^{T_p s}$ and $H(s) = K$, a simple proportional control is composed, and with respect to Kondo's "shaft" driver model in Figure 38.1, the steering angle follows from $\delta_s(s) = Ke^{T_p s} y(s)$ for desired $y_0 = 0$.

The term $v\psi(t)$ in (38.1) was replaced by $\dot{y}(t)$ for small $\psi(t)$ and low frequencies ($s \to 0$); thus, $B(s) = 1 + T_p s$ [22–24], and more often, by applying a Taylor series expansion for $y(t + T_p)$, a second-order prediction function is used (e.g., [19,25–28]):

$$B(s) = 1 + T_p s + \frac{1}{2} T_p^2 s^2 \tag{38.2}$$

Descending from the interest in human tracking and pilot/aircraft control, many contributions between 1959 and 1980 came from McRuer, Krendel, Weir, and others (e.g., [29–33]) and influenced human driver modeling strongly up to now. In [29],

$$H(s) = \frac{K(T_L s + 1)}{(T_I s + 1)(T_N s + 1)} e^{-\tau_r s} \tag{38.3}$$

was proposed for the human tracking behavior (without $B(s)$ and $P(s)$ in Figure 38.2). While the reaction time τ_r and T_N (neuromuscular delay time) are considered as human properties independent of the task, K, T_L, T_I are associated with the controlled system. Later, this approach was further modified to the "precision model" [32] and adopted, for instance, in [16,26] (see also later Equation 38.7).

Already in [2], Kondo introduced a second driver model, which represents a linear feedback of the deviation of yaw angle, course angle, and lateral position, and many following models can be seen as modifications and enhancements of this basic idea.

In [33,34], for instance, a driver model is proposed, which is, in short, a compensatory feedback of the deviation of the lateral position Δy with respect to a reference position and the yaw angle error $\Delta\psi$

$$\delta_s = Y_y Y_\psi \Delta y + Y_\psi \Delta\psi + \frac{K_c}{K_\psi} Y_\psi r_c + n \tag{38.4}$$

but also makes use of the local roadway curvature ρ_r^* (at the car's center of gravity (CG) position) and the vehicle forward speed u in $r_c = \rho_r^* u$; n is usually a filtered noise and represents a driver's remnant. The driver's transfer functions in (38.4) are specified by

$$Y_\psi = K_\psi (T_L s + 1) e^{-\tau_r s}, \quad Y_y = K_y \tag{38.5}$$

In literature [16], this model is often referred to as "STI pursuit driver model" or "STI compensatory driver model" when the roadway is straight or gently curving and ρ_r^* not taken into account and $\Delta\psi = \psi$ in (38.4) [34].

Reid et al. refined the model by integrating preview [16,35]. The control law reads

$$\delta_c = K_c\rho^* + K_\psi \Delta\psi_e + K_y \Delta y_e \tag{38.6}$$

where

- ρ^* is the road curvature previewed at the point on the road ahead of the car that it will reach after the preview time T_p
- $\Delta\psi_e$ is the difference between the vehicle heading angle and the roadway heading angle at a point where a line drawn from the driver's eye meets the lane centerline at a right angle
- Δy_e is the distance from the vehicle's centerline at a point adjacent to the driver's eye to the point where the above line meets the lane centerline

The applied steering angle δ_s to the vehicle follows then from a first-order lead–lag element with time delay τ_r, $\delta_s = H(s)\delta_c(s)$ with (here $K_M = 1$)

$$H(s) = K_M \frac{T_L s + 1}{T_I s + 1} e^{-\tau_r s} \tag{38.7}$$

that is a simplified form of the "precision model" [30,31].

Likewise, Allen et al. [36] consider the more illustrative visual angle between the vehicle's longitudinal axis and the aim (or sight) point, corresponding to $\Delta\psi_c$ in Figure 38.1. The difference ψ_f between $\Delta\psi_c$ and the visual angle ψ_v, that is, the angle between the vehicle's longitudinal axis and the point on the predicted vehicle path (along the current curvature ρ_v^*) close to the aim point

$$\psi_f = \Delta\psi_c - \psi_v, \quad \psi_v \approx \frac{L}{2}\rho_v^* \tag{38.8}$$

respectively the curvature error, $\rho_e^* = 2\psi_f / L$, is then to be eliminated by the driver. An additional feedback is included that accumulates the lateral deviation error Δy_e whenever the vehicle is not correctly positioned laterally on the lane. This visual feedback concept is meant to model the driver's steering control law, without implying that the driver visually focuses on any particular point in the driving scene, but scans actively the whole scene. Finally, the feedback of the yaw rate is intended as a motion feedback cue providing a high-frequency lead that the driver can use to compensate for vehicle lag.

For a recent adaptation and application of this model, see [37–39].

The well-known "crossover model" [30,31,33,40] provides a very effective way for the description of human regulation control tasks and is consequently often used for the parametrization of human driver models. The open-loop transfer function $H(s)G(s)$, which represents the combined system of driver $H(s)$ and vehicle $G(s)$, is described by the experimentally observed closed-loop behavior with

$$H(s)G(s) = \frac{\omega_c}{s} e^{-\tau_r s} \tag{38.9}$$

that means with integral behavior (–20 dB/decade) and adapted phase shift due to the dead time element ($\leq 90°$), in the immediate vicinity of the crossover frequency ω_c. The parameter τ_r is the effective system dead time.

The capability of the human driver for adaptation is achieved on the basis of the invariant overall system behavior; as a consequence the driver is able to compensate tracking errors despite changing

vehicle characteristics. Although the model looks very generally applicable, it faces limitations and the parameter may depend on the kind of input signal. In particular, in [13], it is pointed out that on the basis of driver–vehicle measurements [41], an increased gain (toward −60 dB/decade) and phase shift at low frequencies need to be taken into account for this modeling approach; at mid-range frequencies, the "crossover model" behavior is appropriate just in the immediate vicinity of the crossover frequency. For more recent (simulator) tests and values for crossover frequencies and phase margins at different driving maneuvers, see [27].

Many models are obliged to the aforementioned STI driver model idea; they may also show some modifications (e.g., [23,42]) or lead to rather own models [43]; others are simplifications of the aforementioned ideas (e.g., [44]); nevertheless, these kinds of models are often used as substitutes of the driver up to now.

A driver model that was followed up by others as well (e.g., [26,27,45–48]) is the two-level model from Donges [49]. In Figure 38.3, the two levels (or layers) can be identified, an anticipatory open-loop and a compensatory closed-loop control.

The open-loop control represents the guidance level of steering activity and simulates the anticipatory response to the deterministic run of the desired path curvature. The required parameters of this level are estimated by making use of the measurements of the desired path curvature and the driver's steering wheel angle and optimizing a quadratic criterion; thus, the anticipatory control is formulated in the discrete-time form of an nth-order scalar difference equation with discrete-time points $t_k (k = 1, 2, \ldots)$:

$$\delta_a(t_k) + a_n\delta_a(t_{k-1}) + \cdots + a_1\delta_a(t_{k-n}) = b_n\kappa_r(t_{k-1} + T_A) + \cdots + b_1\kappa_r(t_{k-n} + T_A) \tag{38.10}$$

where

T_A is the anticipation time
κ_r the desired path curvature
a_i, $b_i (i = 1 \ldots n)$ are time-invariant constants

As the differences in order n are small, a first-order approximation with two parameters is suggested, while in [26,50],

$$T_{2s}^2\ddot{\delta}_a(t) + T_{1s}\dot{\delta}_a(t) + \delta_a(t) = V_a\kappa(t + T_a) \tag{38.11}$$

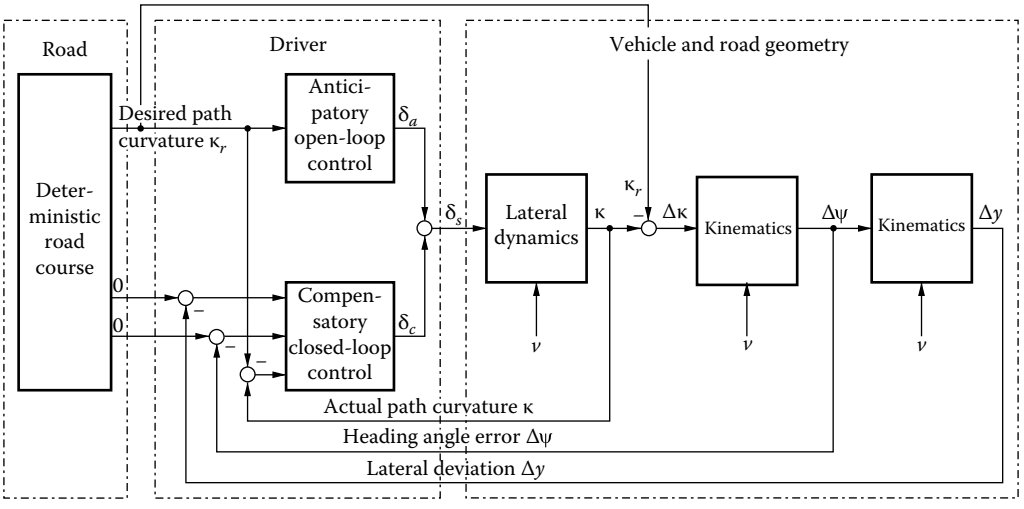


FIGURE 38.3 Two-level model from Donges. (From Donges, E., *Hum. Factors*, 20(6), 691, 1978.)

is proposed for the determination of the open-loop steering angle. The determination of the required steering angle for a two-wheel vehicle model to perfectly follow a given trajectory is used to set the parameters V_a, T_{1s}, T_{2s}, which then just depend on vehicle parameters and vehicle speed. It is thus assumed that the main contribution to the steering angle derives from the experience of the driver on the steering behavior of the vehicle and the steering wheel angle follows a given previewed path curvature at a single point ahead.

The success of the feedforward level has to be controlled in the stabilization level of the driver's steering activity (compensatory closed-loop control). Donges' driver in [49] compares actual and desired path curvature, that is, path curvature error $\Delta\kappa$, and observes heading angle error $\Delta\psi$ and lateral deviation Δy. The mathematical formulation of this compensatory closed-loop control is

$$\delta_c(t) = -\left[h_\kappa \Delta\kappa(t-\tau_r) + h_\psi \Delta\psi(t-\tau_r) + h_y \Delta y(t-\tau_r)\right] \tag{38.12}$$

where

τ_r is the human time delay

h_κ, h_ψ, h_y are gain factors, which are again computed from the difference between the measurements and the feedforward contribution and minimizing a quadratic criterion

They can be chosen to meet the features of the crossover model as well. This is the case in the model of Mitschke [26,27]. In the compensatory control, the predicted deviation $y(t+T_p)$ is the only feedback quantity, which is estimated by applying the transfer function $B(s)$ in (38.2) to $y(s)$. The simplified precision model (38.7) is suggested as a transfer function for the driver.

Because both control levels will not completely reproduce the driver's steering wheel angle, a remnant quantity is a third integral part of the model. The remnant comprises driver-induced signals that are not related to the steering task as well as shortcomings of the special structure and features of the model [49].

In [46], the two-level driver model following Mitschke is extended by a third level that takes also local deviations into account. The local control level is activated in critical situations when the other two levels fail and replaces them for a short period to reduce local deviations directly; a sliding controller is used considering physical limits of the driver. Another enhancement is published in [47]. The driver's internal vehicle model is refined by locally linearizing essential nonlinearities characterizing vehicle handling behavior. In this way, the driver model is enabled to perform good tracking behavior also at higher lateral accelerations. Tracking accuracy is further increased by incorporating information on the change of curvature in clothoids and the local curvature of vehicle motion in the prediction of the anticipated vehicle position. Longitudinal control is added on a corresponding two-level modeling approach in [51].

The multilevel driver model in [52], where all levels contribute to the steering angle, includes besides the three previous (but differently modeled) task levels "steering due to experience" as a fourth level. Adaptation of the driver (and his or her model parameters) to vehicle handling properties is characteristic for the anticipatory level but would be required, in particular at difficult but recurring driving situations, for the second and third level as well. Alternatively, the author suggests to incorporate the perceived mismatch of the expected lateral acceleration and steering moment with the actual values, $a_y(t)$, $M_s(t)$,

$$\delta_e(t) = k_1\left[k_y v^2 \delta_s(t) - a_y(t)\right] + k_2\left[k_M v^2 \delta_s(t) - M_s(t)\right] \tag{38.13}$$

with appropriate (vehicle) constants k_y, k_M and (driver) gains k_1, k_2.

As drivers are looking at a finite interval of the future path ahead, it is assumed for a major class of driver models that the driver is aiming to minimize tracking errors within this interval. Tracking errors can be composed of different kinds of variables, like lateral deviation and

heading angle and usually result from the difference between the previewed and desired path and the predicted vehicle motion.

An often referenced preview/prediction driver model, which is based on optimal control theory, has been presented by MacAdam [53]; it is the basis of the CarSim/TruckSim software package. The linear vehicle model is described in state-space representation by

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \underline{b}\delta_s \tag{38.14}$$

$$y = \underline{c}^T \underline{x} \tag{38.15}$$

with the state vector $\underline{x} = [y, v_y, \psi, \dot{\psi}]^T$, the input steering angle $\delta_s(t)$, and the output lateral displacement of the vehicle mass center $y(t)$. Implying a constant input $\delta_s(t+\tau), \forall \tau \in [0, T_p]$ over the preview interval $[t, t+T_p]$, the output $y(t+\tau)$ of the system (38.14), (38.15) is composed of the zero input and zero-state response by

$$y(t+\tau) = \underline{c}^T \mathrm{e}^{\mathbf{A}\tau}\underline{x}(t) + \underline{c}^T \left(\int_0^\tau \mathrm{e}^{\mathbf{A}\eta} \mathrm{d}\eta \right) b\delta_s(t) \equiv \underline{f}^T(\tau)\underline{x}(t) + g(\tau)\delta_s(t) \tag{38.16}$$

The optimal control $\delta_{s,opt}$, which minimizes the performance function

$$J = \frac{1}{T_p} \int_t^{t+T_p} [y_d(\eta) - y(\eta)]^2 \delta(T_p) \mathrm{d}\eta \tag{38.17}$$

where

$y_d(t)$ is the desired lateral displacement

$y(t)$ is the actual lateral displacement

$\delta(T_p)$ is the Dirac delta function for $[0,T_p]$ and thus utilizes single-point preview, can be derived by substituting (38.16) into (38.17) and setting the partial derivative of J with respect to δ_s zero:

$$\delta_{s,opt}(t) = \frac{y_d(t+T_p) - \underline{f}^T(T_p)\underline{x}(t)}{g(T_p)} \tag{38.18}$$

For a nonoptimal $\delta_s(t)$, the corresponding error $\epsilon(t+T_p) = y_d(t+T_p) - y(t+T_p)$ follows from (38.16) to

$$\epsilon(t+T_p) = y_d(t+T_p) - \underline{f}^T(T_p)\underline{x}(t) - g(T_p)\delta_s(t) \tag{38.19}$$

Substituting now (38.18) in (38.19) yields another expression for the optimal steering angle:

$$\delta_{s,opt}(t) = \delta_s(t) + \frac{\epsilon(t+T_p)}{g(T_p)} \tag{38.20}$$

that is used to take delay limitations of the human driver into account by introducing a transfer function $H(s)$ between $\delta_{s,opt}$ and the actual input δ_s to the vehicle. For $H(s) = \mathrm{e}^{-\tau_r s}$, with the effective transport lag τ_r, and the vector transfer function $\mathbf{G}(s) = [\mathbf{I}s - \mathbf{A}]^{-1}\underline{b}$ for the controlled element, the resulting driver model is displayed in Figure 38.4. The parameters T_p and τ_r can be used for adaptation to measurements and/or crossover model behavior.

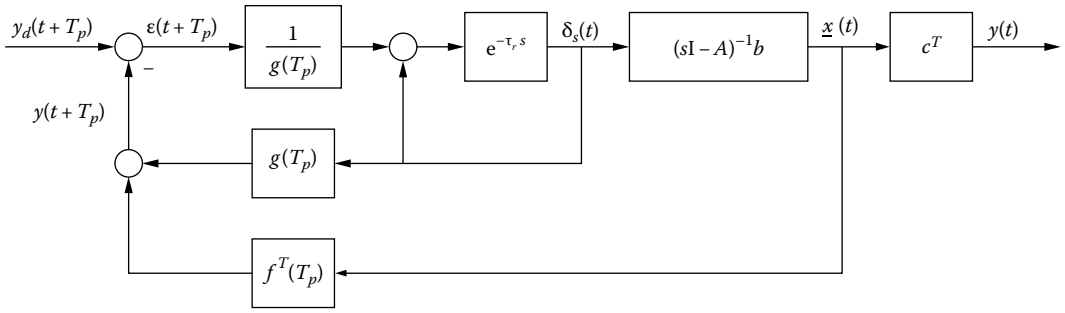


FIGURE 38.4 Optimal (single-point) preview control model. (From MacAdam, C.C., *IEEE Trans. Syst., Man, Cybern.*, SMC-11(6), 393, 1981.)

A recent extension of the previous model is proposed in [54,55]. The performance function to be minimized for the optimal steering angle is changed to

$$J = \int_{t}^{t+T_p} [s(\eta)]^2 d\eta \tag{38.21}$$

where $s(\eta)$ is defined by

$$s(\eta) = \left(k w_{\dot{y}}(\eta) \frac{d}{dt} + w_y(\eta) \right)^{n-1} \Delta y(\eta), \quad \Delta y(\eta) = y(\eta) - y_d(\eta) \tag{38.22}$$

with the weighting parameters $k, 0 \le w_{\dot{y}} \le 1, 0 \le w_y \le 1$. For $n = 2$ in (38.22), $\Delta\dot{y}(t) = \dot{y}(t) - \dot{y}_d(t)$ is considered and can also be interpreted as

$$\Delta\dot{y}(t) = \dot{y}(t) - \dot{y}_d(t) = v_y + v_x(\psi - \theta_d) \tag{38.23}$$

where

v_y is the lateral

v_x is the longitudinal speed and the term $(\psi - \theta_d)$ the vehicle yaw angle error

The gain k in (22) penalizes $\Delta\dot{y}(t)$; that means that drivers with a high gain pay more attention to the vehicle's yaw motion.

In [53], the desired trajectory is specified in the inertial coordinate system but is transformed here to the local (vehicle body) coordinate system to reflect the driver's viewpoint, similar as later shown in Figure 38.5. The derived steering angle may change once at half of the preview time; without this functionality and $k = 0$, the model reduces to MacAdam's driver model.

Another modification is the prediction of the vehicle motion. The driver learns to adapt to a vehicle gradually by learning to "invert the vehicle model" [54]. The learning process is approximated by a least square optimization of a fourth-order auto-regressive integrated moving average with external inputs (ARIMAX) model

$$A(q^{-1})y(k) = B(q^{-1})q^{-d}u(k-1) + \frac{e(k)}{1-q^{-1}} \tag{38.24}$$

with $d = 0$ and the z-domain operator q. $A(q^{-1})$ and $B(q^{-1})$ are recursively identified from rich steering excitations $u(k)$ of a nonlinear vehicle model and the resulting lateral displacement response $y(k)$ by minimizing the sum of squares of the residual error $e(k)$.

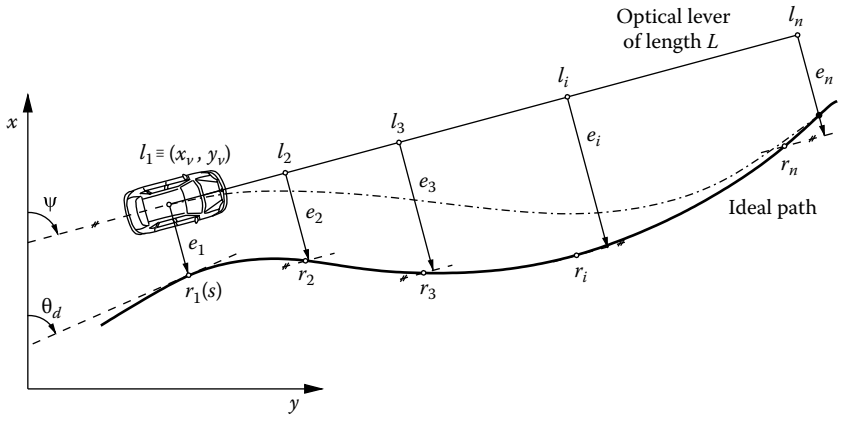


FIGURE 38.5 Preview schema (From Sharp, R.S. et al., *Vehicle Syst. Dyn*., 33, 289, 2000); the line-dotted trajectory refers to. (From Frezza, R. et al., *Proc. 2003 IEEE/ASME Int. Conf. Adv. Intell. Mechatron*., 1, 36, 2003.)

In [19], it is assumed that drivers get their basic experience from quasi-static driving, where the lateral acceleration of the vehicle (respectively the curvature of the vehicle trajectory) is about proportional to the steering angle input. That is why a performance criterion similar to (38.17) is minimized with respect to the lateral acceleration, which is then further processed by a transfer function according to (38.7) to derive the driver's steering angle. Single-point preview is applied, the lateral motion of the vehicle predicted using (38.2). Further, it is suggested to compose the transfer function of the driver/vehicle closed loop of a preview transfer function $P(s)$, describing the driver's preview ability, and a follower transfer function $F(s)$, which reflects the driver's following process. The authors propose $y(s)/y_d(s) = P(s) \cdot F(s) \approx 1$ for low frequencies as a control design rule, with the lateral position y and its desired value y_d. Longitudinal control is added in [56]. The method is further generalized in [57] for lateral control and also extended by longitudinal control in [58].

The concept of multipoint preview is clearly presented by Sharp et al. [59] (see Figure 38.5); a similar idea has been introduced in [60].

At position s along the trajectory, the driver experiences the right-angled deviations e_i between a set of n points $l_1, l_2, \ldots, l_n$ on an optical lever and corresponding points $r_1, r_2, \ldots, r_n$ on the desired path. The length of the optical lever is $L = T_p v_x$ (with longitudinal velocity v_x); the points can be equally spaced or more concentrated toward the CG of the vehicle, as realized in [59] with eight points. The distance $\overline{l_i l_{i+1}}$ is the same as distance $\overline{r_i r_{i+1}}$, and thus, the points $r_1, r_2, \ldots, r_n$ do not lie exactly on the direction perpendicular to the optical lever through the $l_1, l_2, \ldots, l_n$ points depending on the complexity of the trajectory geometry. It is reported that these errors seem not to have any significant influence on the robustness of the driver. The first point l_1 coincides with the CG of the vehicle and e_1 represents the vehicle lateral offset from the current intended position on the path. Besides the state feedback e_1, the difference $(\psi - \theta_d)$ between the vehicle yaw angle and the tangent angle of the intended path at r_1 is another state feedback used for the driver's steering control. The steering angle is then a linear combination of the control inputs

$$\delta_s = K_\psi(\psi - \theta_d) + K_1 e_1 + \sum_{i=2}^{n} K_i e_i \tag{38.25}$$

with the control gains $K_\psi, K_1, \ldots, K_n$. Although no systematic determination of these parameters is done, some general rules are given, for example, that the driver model preview gain sequence should be exponentially convergent to zero toward distant preview samples. Further, for extreme

driving conditions, saturation functions are applied on the weighted inputs in (38.25) to account for unreasonable high steering angles resulting in tire forces far beyond saturation.

This driver model also controls the longitudinal vehicle motion, even though completely uncoupled from the lateral control, and no human driver specifications are included—typical for many applied velocity tracking controllers. An ideal velocity profile needs to be provided, and with actual and target velocity a short distance ahead, a required constant acceleration a is evaluated. The longitudinal force F to maintain the desired forward motion is then

$$F = ma + I_f \frac{a}{r_f^2} + (I_r + I_e i_g^2)\frac{a}{r_r^2} + \frac{1}{2}\rho A c_x v_x^2 \tag{38.26}$$

with vehicle mass m and the moments of inertia I_f, I_r, I_e of front and rear axle rotating masses and inertia due to engine, which depends on the gear ratio i_g and the aerodynamic drag (aerodynamic drag coefficient c_x, longitudinal velocity v_x); drag by the steering wheels at cornering is neglected. When F is positive, a driving torque $T = Fr_r$ (rear wheel radius r_r); otherwise, an engine braking torque is applied, and a further residual braking torque is split to the front and rear axle according to a fixed braking ratio.

The (lateral) model was developed further by Sharp et al. in [61,62] by introducing the linear-quadratic regulator (LQR) method for the determination of the preview control parameters. The state-space representation ((38.14), (38.15)) of the two-wheel vehicle model is transferred to a time-discrete formulation, that is, the first row in (38.27):

$$\underline{z}(k+1) = \begin{bmatrix} \underline{x}(k+1) \\ \underline{y}_r(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_v & 0 \\ 0 & \mathbf{D} \end{bmatrix} \begin{bmatrix} \underline{x}(k) \\ \underline{y}_r(k) \end{bmatrix} + \begin{bmatrix} \underline{0} \\ \underline{e} \end{bmatrix} y_{ri}(k) + \begin{bmatrix} \underline{b}_v \\ \underline{0} \end{bmatrix} \delta_s(k) \tag{38.27}$$

$$\underline{z}(k+1) = \mathbf{A}_c \underline{z}(k) + \underline{e}_c y_{ri}(k) + \underline{b}_c \delta_s(k) \tag{38.28}$$

The previewed road is described by the absolute lateral displacements in $y_r(k) = [y_{r0}, y_{r1}, \ldots, y_{rn}]^T$; thus, the second row of (38.27) is a shift process with respective $\mathbf{D}$ and $\underline{e}$ and the road input y_{ri} that becomes y_{rn} at the next step. The preview horizon is again n and $\underline{z} = [y, \dot{y}, \psi, \dot{\psi}, y_{r0}, y_{r1}, \ldots, y_{rn}]^T$.

The time-invariant LQR control is found from minimizing the quadratic cost function under white noise excitation:

$$J = \lim_{j\to\infty} \sum_{k=0}^{j} \left[\underline{z}^T(k) \mathbf{R}_1 \underline{z}(k) + \delta_s(k) r_2 \delta_s(k) \right] \tag{38.29}$$

with the weights $\mathbf{Q}$ in $\mathbf{R}_1 = \mathbf{C}^{\mathrm{T}}\mathbf{Q}\mathbf{C}$ and r_2. With sampling time T, respectively, sampling distance vT and by defining

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & 1/(vT) & -1/(vT) & 0 & 0 & 0 & \ldots & 0 \end{bmatrix} \tag{38.30}$$

not only the squares of the differences between the y-coordinate of the car's predicted reference point and the corresponding value y_{rj} for the lateral displacement of the road are minimized but also the squares of yaw angle differences $\{\psi(k) - [y_{r1}(k) - y_{r0}(k)]/(vT)\}$. After solving the algebraic Riccati equation

$$\mathbf{P}_\infty = \mathbf{A}_c^T \mathbf{P}_\infty \mathbf{A}_c - \mathbf{A}_c^T \mathbf{P}_\infty \underline{b}_c \left(\underline{b}_c^T \mathbf{P}_\infty \underline{b}_c + r_2 \right)^{-1} \underline{b}_c^T \mathbf{P}_\infty \mathbf{A}_c + \mathbf{C}^T \mathbf{Q} \mathbf{C} \tag{38.31}$$

the optimal steering angle results in

$$\delta_{s,opt}(k) = -\mathbf{K}_\infty \underline{z}(k) \quad \text{with } \mathbf{K}_\infty = \left(\underline{b}_c^T \mathbf{P}_\infty \underline{b}_c + r_2 \right)^{-1} \underline{b}_c \mathbf{P}_\infty \mathbf{A}_c \tag{38.32}$$

This preview control is based on a fixed reference system, as the preview driver model of MacAdam in [53], while in reality, all path observations are made by the driver from his or her moving position, and consequently, the authors in [61] transform the gain with respect to a relative system, similar to Figure 38.5.

Although the three driver models in [53,54] and [61,62] are very similar and rely all on optimal control theory (predictive control and LQR), they differ in some aspects: the applied performance criterion, use of absolute and relative lateral path and/or yaw angle information, preview horizon, and for how long the steering angle can change within the preview horizon. A thorough comparison and analysis of the differences of the aforementioned three approaches is done in [63].

The driver model in the commercial MBS software Adams/Car is based on [64]. The driver is organized in three hierarchical levels. The task planner sets the goal of the controller; the output is an optimal trajectory based on a criterion that mathematically codes the given task. The tactical planner is explained in more detail (see also [65]). This level computes an optimal trajectory based on the current state of the vehicle relative to the ideal trajectory. Trajectories are optimal in the sense that they fit at best the road in the interval $[t, t + T_c]$, where T_c is called the control horizon, equivalent to L in Figure 38.5, where an optimal trajectory (dash-dotted line) is illustrated; the ideal trajectory is also measured at n look ahead distances, although here exactly on the direction perpendicular to the longitudinal vehicle axis through the points $l_1, l_2, \ldots, l_n$.

For each time step t, the desired path is described by a function $y = p(x) = \sum_{j=0}^{r} a_j x^j$ in the vehicle body frame coordinates. For the order of the polynomial, which has to meet the initial conditions $p(0) = 0$, $\mathrm{d}p/\mathrm{d}x = 0$, $r = 4$ is suggested. The coefficients are then computed with minimum square fit taking into account weights on the preview distances l_i and may depend on speed v and curvature κ constraints (bounded steering angles). A "kinematic vehicle model" with wheelbase $l_a + l_b$ is applied:

$$\dot{x} = v\cos\psi, \quad \dot{y} = v\sin\psi, \quad \dot{\psi} = \frac{\tan\delta_s}{l_a + l_b} v = \kappa v \tag{38.33}$$

and no dynamics and no slip are considered in this level. From (38.33), there follows $\dot{\psi} = \kappa v$ and thus the desired yaw rate

$$\dot{\psi}_d = \left.\frac{d^2 p(x)}{\mathrm{d}x^2}\right|_{x=0} v = 2a_2 v \tag{38.34}$$

which is to be set in the third level. The motion controller acts on the throttle, brake, and clutch to track a given velocity profile as well as on the steering wheel. By applying the two-wheel vehicle model (with moment of inertia I_z), assuming that the front lateral tire force F_{yf} can be controlled directly and by setting

$$\ddot{\psi} = \frac{1}{I_z}(l_a F_{yf}\cos\delta_s - l_b F_{yr}) \equiv \tilde{u} \tag{38.35}$$

where the rear lateral tire force F_{yr} is determined kinematically, the control $\tilde{u}$ is derived to achieve $\dot{\psi} = \dot{\psi}_d$ by simply designing a tracking controller for a plant whose transfer function is an integrator. With known $F_{yf} = F_{yf}(\alpha_f)$, the steering angle results finally from the estimated front sideslip angle α_f.

An example for using nonlinear control to represent a driver is [66–68], and the idea behind is applied in the driver model of the commercial software veDYNA [69]. A guidance level sets the desired trajectory and velocity profile (see also Section 3.2 [68,70]), in particular position and velocity of a "target point". In the stabilization level, a position controller tries to match a "target aim point" with the current "target point"; the "target aim point" is fixed on the vehicle's longitudinal

axis a specified preview distance ahead, although the CG of the vehicle with position x, y is chosen in [66,67] and in the following. The output of the position controller is a suitable steering wheel angle and brake/accelerator pedal positions to minimize the deviations from the target path and given or calculated speed profile.

On the basis of a nonlinear extended two-wheel vehicle model in state-space representation,

$$\underline{\dot{x}} = \underline{a}(\underline{x}) + \mathbf{B}\underline{u} \tag{38.36}$$

$$\underline{y} = \underline{c}(\underline{x}) \tag{38.37}$$

$$\underline{x} = [\beta, \psi, \dot{\psi}, v, x, y]^T, \quad \underline{y} = [x, y]^T, \quad \underline{u} = [F_{yf}, F_x]^T \tag{38.38}$$

the position x, y is controlled by deriving appropriate front lateral tire forces F_{yf} and longitudinal tire forces F_x; in this way, a combined control of longitudinal and lateral dynamics by the driver is achieved instead of a separated view. The nonlinearities in $\underline{a}(\underline{x})$ are mainly due to $\dot{x}$, $\dot{y}$, which depend nonlinearly on velocity v, yaw angle ψ, and sideslip angle β of the vehicle, details in [66,67]. Following the nonlinear decoupling method, the output equations in (38.37) have to be differentiated as often as at least one input that appears explicitly, in this case each of them two times. The second derivatives of the output vector $\underline{y}^*(t) = [\ddot{y}_1(t), \ddot{y}_2(t)]^T$ can now be expressed by

$$\underline{y}^*(t) = \underline{c}^*(\underline{x}(t)) + \mathbf{D}^*(\underline{x}(t))\underline{u}(t) \tag{38.39}$$

and respective matrices $\underline{c}^*$, $\mathbf{D}^*$ easily derived with (38.36) through (38.38). Considering (38.39) and presupposing that $\mathbf{D}^{*-1}(\underline{x}(t))$ exists, the control law reads

$$\underline{u}(t) = \mathbf{D}^{*-1}(\underline{x}(t))(-\underline{c}^*(\underline{x}(t)) + \Lambda\underline{w}(t) - \underline{q}^*(\underline{y}(t))) \tag{38.40}$$

where $\Lambda = \text{diag}\,(\lambda_i)$ is a weighting matrix for the desired position in $\underline{w}(t)$ ("target point"). By choosing

$$q_i^*(\underline{y}(t)) = \alpha_{1i} y_i(t) + \alpha_{2i}\dot{y}_i(t), \quad i = 1, 2 \tag{38.41}$$

and substituting (38.40) in (38.39), the dynamics of the system

$$\ddot{y}_1(t) + \alpha_{21}\dot{y}_1(t) + \alpha_{11} y_1(t) = \lambda_1 w_1(t) \tag{38.42}$$

$$\ddot{y}_2(t) + \alpha_{22}\dot{y}_2(t) + \alpha_{12} y_2(t) = \lambda_2 w_2(t), \tag{38.43}$$

can be adjusted with the constants λ_i and α_{1i},α_{2i} ($i = 1, 2$) to guarantee $\underline{y}(t) \to \underline{w}(t)$ in a linear characteristic of second order. With the resulting $u(t)$ and front lateral tire force $F_{yf} = F_{yf}(\alpha_f)$, the sideslip angle α_f and consequently the steering angle are determined; the control F_x is transferred to adequate brake/accelerator pedal positions.

Another type of nonlinear controller used for representing the driver is FLC. Fuzzy logic interfaces are said to resemble human thinking and the process of compromise in making a decision [71] and are in particular useful when mathematical models are very complicated. FLC offers the ability to develop rules that make intuitive sense and can be expressed in linguistic terms [72]. Nevertheless, there are fewer driver models with FLC to be found in literature than with previous classical control theory approaches, at least when models for virtual test drivers are considered. Two driver models are presented here as representatives in more detail ([72] and [73]); others are listed thereafter or in subsequent sections.

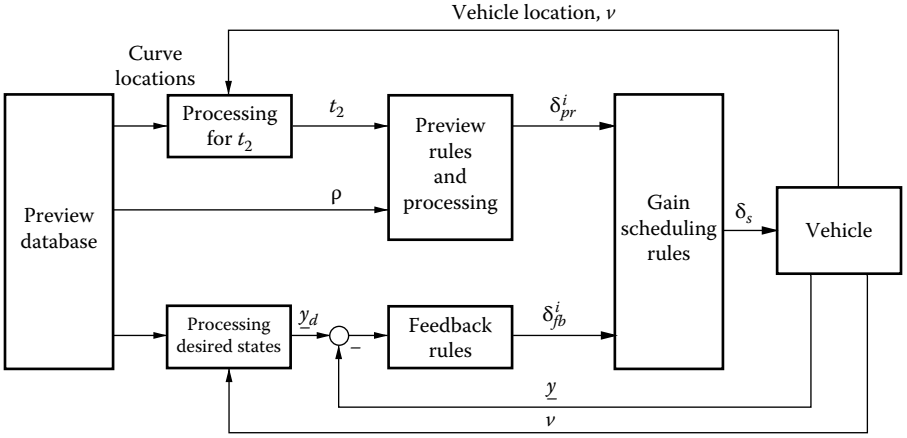


FIGURE 38.6 Driver model with FLC. (From Hessburg, T. and Tomizuka, M., *Contr. Syst. Mag.*, 14, 55, 1994.)

The block diagram of the feedback control in [72] is plotted in Figure 38.6. It is assumed that preview information regarding location, direction, and radius of curvature for each curved section on the reference track can be supplied. The controller is modularized in feedback, preview, and gain scheduling rules.

The input vector for the feedback rules is $\underline{y} = [y, \dot{y}, (\dot{\psi} - \dot{\psi}_d)]^T$, where y denotes the lateral error with respect to the road centerline and $(\dot{\psi} - \dot{\psi}_d)$ the rate of the heading angle error (see Figure 38.7a). The interference rules with linguistic variables *LE* (corr. to y), *CLE* (corr. to $\dot{y}$), *YWR* (corr. to $(\dot{\psi} - \dot{\psi}_d)$), and *DFB* (corr. to δ_{fb}) take the form

$$\text{IF } LE \text{ IS } A_{LE} \text{ AND } CLE \text{ IS } A_{CLE} \text{ AND } YWR \text{ IS } A_{YWR} \text{ THEN } DFB \text{ IS } A_{DFB} \tag{38.44}$$

with 5 different subsets each ("neg. big", "neg. small", "nil", "pos. small", "pos. big") corresponding to A_{LE}, A_{CLE}, and A_{YWR}; typical triangular membership functions are used [72], and common min–max method is applied. Consequently, for defuzzification, singletons (for now 9 different subsets, "neg. huge"…"nil"…"pos. huge", corresponding to A_{DFB}) are introduced and a resulting crisp steering angle δ_{fb} is obtained.

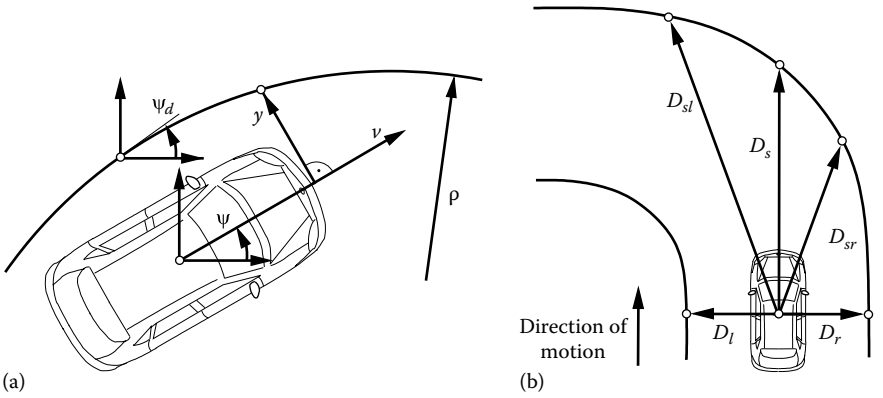


FIGURE 38.7 Sensing information for FLC driver model, (a) (From Hessburg, T. and Tomizuka, M., *Control Syst. Mag.*, 14, 55, 1994) and (b) (From Zeyada, Y. et al., Driver modeling using fuzzy logic controls for human-in-the-loop vehicle simulations, in: *Current Advances in Mechanical Design and Production VII*, Hassan, M.F., Megahed, S.M., Eds., Elsevier Science Ltd., Oxford, U.K., 2000).

The steering angle δ_{pr} follows the previewed road geometry ahead of the vehicle. The right amount of steering action for "look ahead" control depends on the predicted response characteristics of the vehicle, therefore on parameters p_c and p_n in

$$\delta_{pr} = \frac{p_c}{\rho_c} + \frac{p_n}{\rho_n} \tag{38.45}$$

which describes the preview term for the front wheel steering angle; ρ_c and ρ_n are current and next radii of curvature, respectively. p_c and p_n depend on $\min(t_2,t_w)$, where t_2 is the estimated time for the vehicle to reach the next curve transition and $t_w = 1$ s. With the preview database and velocity v, time t_2 can be estimated. Twenty-six fuzzy subsets to describe t_2 and 52 singletons for p_c and p_n are chosen to accurately (and smoothly) approximate the preview steering action δ_{pr} (at step changes in the radius of curvature).

The gain scheduling rule base is designed to infer the final front wheel steering angle command δ_s from v to account for more gentle steering commands at higher speed. Both the feedback and preview rule bases are designed at three velocities $v_1 = 5$ m/s, $v_2 = 12.5$ m/s, $v_3 = 20$ m/s. Thus, the rule

$$\text{IF } V \text{ IS } A_V \text{ THEN } \delta_s = \delta^i_{fb} + \delta^i_{pr}, \quad A_V \epsilon \left\{ SMALL, MED, HIGH \right\} \tag{38.46}$$

schedules the gains between the parameterizations at v_i ($i = 1,2,3$) and the corresponding $\delta^i_{fb}, \delta^i_{pr}$ for the current velocity v; for membership functions, see [72].

The FLC driver model in [73] not only performs steering inputs but also provides braking events. The model is trained to both curve following and collision avoidance.

Compared to [72], different input information is used in [73] (see Figure 38.7a,b), in particular the left, right, straight-ahead, straight-left, and straight-right distances (D_l,D_r,D_s,D_{sl},D_{sr}) from the CG of the vehicle to the roadway borders. Two FLC modules control steering and braking. The amount of steering necessary to guide the vehicle consists of the contribution from preview, resulting from the anticipation of the incoming curve, and may be described by

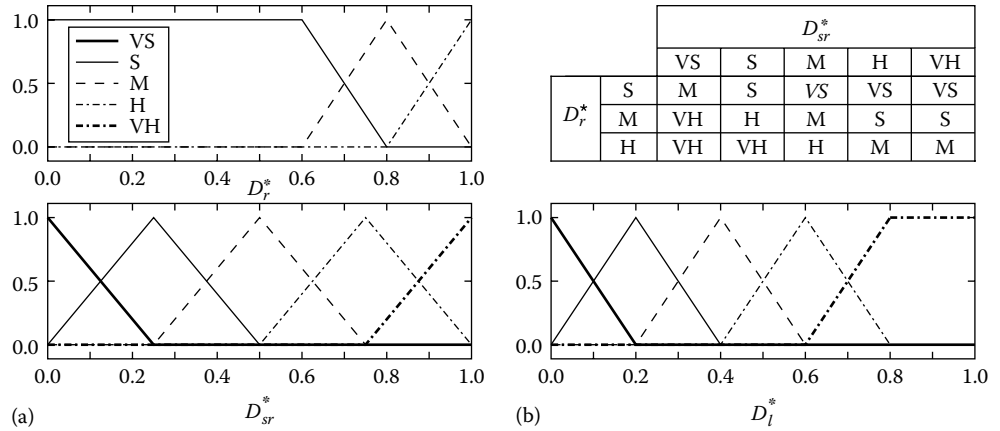
$$\text{IF } \begin{matrix} D_{sl} > D_{sr} \\ D_{sr} > D_{sl} \end{matrix}, \quad \text{THEN } \begin{matrix} \text{left} \\ \text{right} \end{matrix} \text{ steering is required} \tag{38.47}$$

and of the contribution of the control of the deviation of the CG of the vehicle from the roadway border

$$\text{If } \begin{matrix} D_l > D_r \\ D_r > D_l \end{matrix}, \quad \text{then } \begin{matrix} \text{left} \\ \text{right} \end{matrix} \text{ steering is required} \tag{38.48}$$

The steering module consists of the left and right steering part. For the normalized inputs that cause left steering, $D^*_r = 1 - 2D_r/D_w$ and $D^*_{sr} = D_{sr}/D_{sr,max}$ (where D_w is the road width and $D_{sr,max}$ the maximum allowable value for D_{sr}), the membership functions are shown in Figure 38.8a. Following the interference rules in the same figure (right, top) with five different subsets ("very small", "small", "med", "high", "very high") for D^*_l a crisp value δ_l for the left steering part can be assigned by applying the membership functions (right, bottom), with $D^*_l = \delta_l/\delta_{max}$ (where δ_{max} is the maximum allowable steering angle).

The right steering part is handled analogously. For running straight-ahead on the road centerline, the steering angles from both parts equalize each other. It should be mentioned that no steering output can be associated with low values of D^*_r in this way and what should simulate the human driver behavior of tolerating a small shift from the road centerline.



D_r^* \ D_{sr}^*	VS	S	M	H	VH
S	M	S	*VS*	VS	VS
M	VH	H	M	S	S
H	VH	VH	H	M	M

FIGURE 38.8 (a) Membership functions and (b) fuzzy interference rules for left steering module of FLC driver. (From Zeyada, Y. et al., Driver modeling using fuzzy logic controls for human-in-the-loop vehicle simulations, in: *Current Advances in Mechanical Design and Production VII*, Hassan, M.F. and Megahed, S.M., Eds., Elsevier Science Ltd., Oxford, U.K., 2000.)

For simplicity, the braking module uses only one sensing variable, that is, normalized $D_s^* = D_s / D_{st}$. D_{st} is the stopping distance from the initial velocity v_0 at the time of brake application to a complete stop by maximum braking and is approximated by $D_{st} = v_0/(2a_{dec})$; the assumed constant maximum deceleration is set to $a_{dec} = 0.8g$. The corresponding membership functions with five different subsets for D_s^*, simple interference rules, and five membership functions for the output, the normalized braking command, are omitted here (see [73]). The FLC of the braking model is parameterized in a way that the braking system will respond only for values of D_s less than or equal to three times the expected stopping distance at the current vehicle speed v_0.

In order to get precise fuzzy reasoning rules, a self-organizing method is proposed for the FLC driver model in [74]. The input variables for the fuzzy controller are yaw velocity and lateral position of the vehicle with respect to the desired maneuvers, single- and double-lane change. Normal distribution shapes are considered for membership functions, which are determined by a fuzzy statistical approach (by interviewing many drivers). Each fuzzy variable is denoted by 7 linguistic fuzzy variables, and thus, a total of 49 rules are required, which are then refined by a self-organizing method. Therefore, the 7 linguistic values for the three fuzzy variables, C for steering angle, Y for lateral displacement, and R yaw rate, are denoted by –3, –2, and –1, 0, 1, 2, 3, respectively. To adjust the 49 rules, a revising factor α is introduced in the rule:

$$C = -\langle \alpha Y + (1-\alpha)R \rangle,\ 0 \le \alpha \le 1 \tag{38.49}$$

where $\langle x \rangle$ represents a minimum integer whose absolute value is larger than or equal to x and whose *sign* is the same as x. Computing a set of appropriate α associated with the 49 fuzzy rules provides an ideal fuzzy reasoning rule base.

The lateral component of the FLC driver model in [12] is similar to the aforementioned one [74], as the used input variables are the same. This driver model includes also a longitudinal FLC, which is explained here in more detail as an example for a more humanlike longitudinal motion control compared to the often applied simple velocity controllers. The task of the driver is defined to drive as fast as possible through a given course, which is approximated by straight lines and segments of circles. The strategy, which is further the basis for the fuzzy rules, is in a linguistic rule formulation:

- Brake before the curve as late as possible with maximum braking force, in order to attain maximum possible velocity (determined by radius and arc angle of curve) when entering the curve
- Drive through the curve with maximum velocity (determined by radius and rest of arc angle of curve)

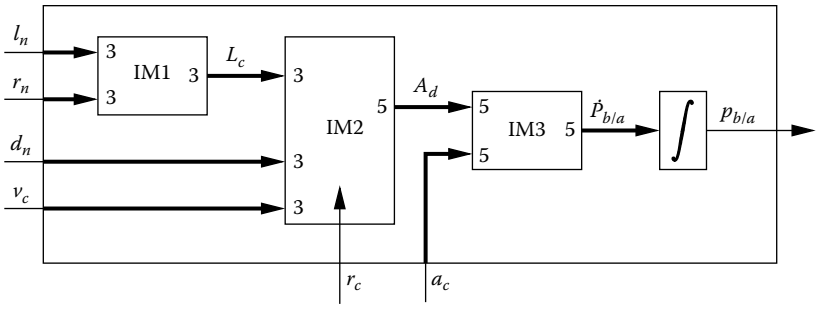


FIGURE 38.9 Longitudinal FLC driver model (From Jürgensohn, T., *Hybride Fahrermodelle*, Pro Universitate Verlag, Berlin, Germany, 1997, ISBN 3-932490-22-3 (in German)); thin lines represent crisp variables, and thick lines linguistic variables; numbers at the input/output of Mamdani interference machine IM (including trapezoidal membership functions) correspond to the number of fuzzy subsets.

- Accelerate at the beginning of a section with maximum possible acceleration until the maximum possible velocity is achieved
- Consider subsequent short curves as a unit

and constitutes the required input variables for the longitudinal FLC in Figure 38.9: arc angle length l_n and radius r_n of the next curve, distance d_n to next section, current velocity v_c and acceleration a_c, and radius r_c of the current curve. The fuzzy output of interference machine IM1 is curve length L_c that denotes an internal, "subjective" variable and is introduced to take into account that drivers can go through short curves with higher speed within the lane width than through long curves of same radius.

As the actuation of the brake or accelerator pedal directly affects the acceleration of the vehicle, the author considers the introduction of an internal desired acceleration A_d, psychologically and physiologically motivated. The current acceleration a_c is then processed to IM3 to follow the internal desired acceleration A_d. Longitudinal acceleration can be (haptically) quickly perceived and controlled by the driver. Therefore, the internal control quantity is the change of the pedal position with respect to time, $\dot{P}_{b/a}$, and consequently, IM3 generates no absolute position but a change, as the driver does when he or she wants to change speed. Defuzzification and integration yield $p_{b/a}$.

Another driver model using FLC and utilizing longitudinal and lateral control is [75] but is referred to Section 38.3.2 for its "course decision process" module. Further FLC driver models are, for instance, shown in [76,77].

The longitudinal virtual test driver in [78] is based on artificial NNs to represent human driving behavior. It is the task of the model to follow a given velocity profile, in this case standardized drive cycles. A conventional feedback controller is responsible for appropriate braking commands, while the throttle command is the output $u(t)$ of two NN controllers (see Figure 38.10).

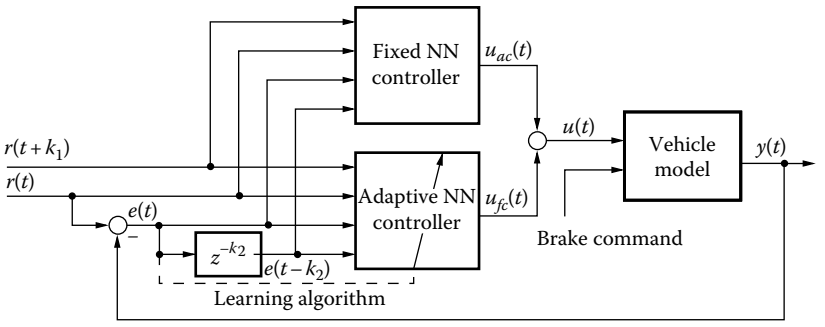


FIGURE 38.10 Longitudinal virtual test driver. (From James, D.J.G. et al., *Proc. Artif. Life Robot.*, 7, 170, 2002.)

The objective of the fixed (parameter) forward NN is to model a human driver following the reference velocity $r(t)$. This is achieved by training the NN with input/target data collected during experimental tests including a human driver. Considering the error $e(t) = r(t) - y(t)$, where $y(t)$ is the velocity of the vehicle, four inputs to the NN are chosen: $e(t)$, $r(t)$, previewed $r(t + k_1)$, and time-delayed $e(t - k_2)$ with $k_1 = 1.1$ s, $k_2 = 0.4$ s. The training process takes into account that both accelerator and brake pedal should not be applied at the same time. The adaptive NN controller is incorporated to enable the driver model to adapt to changes in the vehicle and/or road conditions. This controller is continuously updated to compensate for speed errors $e(t)$. Therefore, $e(t)$ is used in the cost function to optimize the weights and biases of the NN adopting standard backpropagation training. The NN architecture can be seen in Figure 38.12 (with hyperbolic tangent activation function F_1, linear F_2), where more on using NN for driver models can be found in that context.

Besides "transfer function," (optimal) state-space, FLC, and NN modeling methods for setting up a virtual driver, there are still more methodical alternatives, for example, identification methods.

The parameters of an autoregressive moving average with external inputs (ARMAX) model, incorporating structured uncertainties to account for driver model uncertainty due to variation across different drivers, are identified using data from driving simulator tests ([79], also [80]); for the same task, but applying model reference adaptive control (MRAC), see [81]; for online estimation of the ARMAX model parameters, see [82]. Hybrid models are considered in the subsequent sections.

Rather autonomous approaches for driver models are, for example, [83], or the driver model in [84] on a flow chart basis.

Most of the aforementioned driver models apply for a wide variety of applications and have often been introduced in literature rather with exemplary applications. To conclude this section, some implementation examples of mentioned or similar virtual test driver models with respect to the previously stated areas of application are given in [38,39,85–90].

38.3 APPLICATIONS WITH FOCUS ON THE DRIVER

This section addresses driver models and their applications whose interest and concern are mainly related to typical driver tasks and who claim to contribute to a better understanding of (individual) driver behavior.

38.3.1 Understanding Driver and (Individual) Driving Behavior

As already mentioned before, the proposed classification of driver models to specific sections is not distinct and models from other sections may be allocated here as well. In particular, driver models based on the artificial NN approach are qualified for here, as they are often considered to represent adaptive human control behavior and enable the modeling of specific driver/vehicle control interactions. A review of applications of NNs in vehicle systems in general is given in [71]. Although autonomously (humanlike) guided vehicles are not addressed here, synergetic effects may be interesting in this context as well, for example, in [91–99] (or [100–102] with FLC).

Since about 1990, the capabilities of NNs to emulate the human driving behavior have been investigated [103]. Directional control as well as path planning strategies have been modeled by a single-layer functional-link NN (to improve training time and error) and a two-layer feedforward NN, respectively. In [104], a driving assistant system to support overtaking is presented. Input parameters of the system are distances to collision on the left and right lanes and the lateral position and speed of the vehicle. The driver and/or the support system decides on the appropriate accelerator/brake pedal position and steering wheel angle. When the driver takes over the control, a comparison of the decisions is displayed (driver assistant). The signals can be used to teach the NN support system to evaluate the performance of the driver, to teach novices as a trainer, or maybe even to fully substitute the driver (autopilot). A combination of functional-link

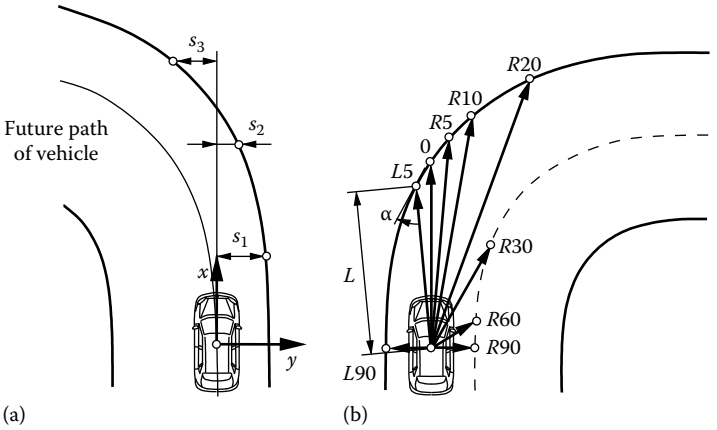


FIGURE 38.11 Radar sensor measurements (a: From MacAdam, Ch.C. and Johnson, G.E., *Vehicle Syst. Dyn.*, 25, 3, 1996) (b: From Kageyama, I. et al., *Vehicle Syst. Dyn. Suppl.*, 33, 122, 1999).

and backpropagation network is preferred to achieve a good compromise between fast-learning and accurate decision-making characteristics.

In [105], an NN-based driver model is both trained by a simple driver model that maps the steering angle as a function of the time-delayed lateral deviation and heading angle and by a human driver in a simulator environment. The input to the NN is the lateral deviation from the desired course and is derived from a simplified view at four points 20, 40, 60, and 80 m ahead of the vehicle; four views are memorized at each time step resulting in 16 data points for the complete input. Required maneuvers for teaching the NNs are analyzed and accuracy of the NNs is compared.

The goal in [106] is the identification of driver steering behavior based upon NNs. The relative lateral displacements s_i ($i = 1, 2, 3$) between the forward projection of the vehicle's longitudinal body axis and the road border (see Figure 38.11a) represent the previewed sensor inputs at time t. One sensor is placed very near to the front end of the vehicle, and the remaining two samples are located at two points ahead of the vehicle. By this specific choice of sensor location and measurements, information about vehicle motion and orientation with respect to the roadway is provided to the NN. To extract further time derivative information from the same basic sensor measurements, time-delayed versions of these signals are utilized. Thus, the augmented set of network inputs, which include $\{s_1(t), s_2(t), s_3(t)\}$ and two corresponding sets of their time-delayed responses $\{s_1(t-\tau_1), s_2(t-\tau_1), s_3(t-\tau_1)\}$, $\{s_1(t-\tau_2), s_2(t-\tau_2), s_3(t-\tau_2)\}$ with $0.1 \le \tau_1 \le \tau_2 \le 0.6$, contains also the key feedback cues required by "classical" driver models for path regulation.

A basic approach to represent nonlinear system behavior for one output is the two-layer NN shown in Figure 38.12. M input signals s_k are processed through weighting values w_{kj}^1 ($k = 1 \ldots M$, $j = 1 \ldots N$) and nonlinear functions F_1 to y_j, which are processed further through weighting values w_{j1}^2 and a linear function F_2 to the final network output z; c_j^1, c_1^2 denote constant bias. In [106], $M = 9$ input signals are applied, further $N = 3$ is chosen, and the sigmoid function $f(u)$ in (38.50) represents the nonlinear contribution of F_1. The network is trained with the backpropagation method, where the training data are obtained from a driver/vehicle numerical simulation, a basic (desktop) driving simulator, and experimental on-road tests.

A similar goal of investigation is followed up in [107], although for different maneuvers (kinds of lane change and slalom, sinusoidal maneuvers). Therefore, the applied discrete-time input vector to the NN is composed of the yaw angular velocity, roll angle and angular velocity, lateral displacement, velocity and acceleration, and a preview lateral offset; the output includes steering angle and steering angular velocity. Different NN architectures are compared, and it is reported that the (Gaussian) radial basis function network (RBFN) showed the best performance with respect to training time and accuracy.

The NN architecture presented in Figure 38.12 is also used to identify driver longitudinal control by representing driver manual control of throttle position during headway keeping tasks [108].

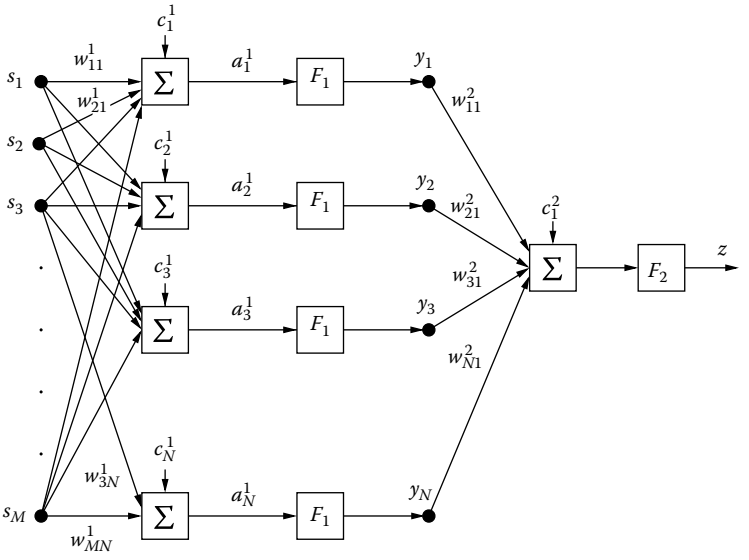


FIGURE 38.12 Two-layer NN architecture (for a single output).

Here, the distance between host and lead vehicle and its time derivative (and their time-delayed counterparts) are combined in the NN to the output for the throttle position. As a kind of adaptive control modeling technique, the NN can then have utility within the context of warning/control packages that require a continuously updated characterization of driver behavior.

In the same paper [108], a similar NN representation of the driver is analyzed to classify specific families of on-highway longitudinal control behavior based on different levels of displayed aggressiveness, respectively, the willingness to pass, follow, or be passed by other vehicles. For this purpose, the pattern recognition ability of certain NN architectures is utilized to automatically identify and classify observed patterns into appropriately defined driving-style categories.

The relationship between human factor and cause of accidents is examined in [109] by using a driver model based on NN. Specific curves, where many accidents are reported, are analyzed with respect to the acquisition of environmental recognition and level of practice to the course. The structure of the applied NN with one hidden layer (of 30 neurons) is similar to the one in Figure 38.12 but with three outputs, in particular steering torque, valve opening of throttle, and braking force. The information used for recognition of the course is derived from the relative position of the environment to the vehicle position using a radar sight description (see Figure 38.11b). The position of the vehicle is mapped by 13 measured lengths to the lane edge L and 13 corresponding angles α, to predict the road shape ahead. Besides these 26 inputs to the NN, further 68 inputs mapping the vehicle behavior are processed. These are 17 current state variables (angles, angular velocities, and angular acceleration for yawing, rolling, pitching, and steering; the velocities and accelerations for the longitudinal and lateral directions; and the sideslip angle of the vehicle), and the same variables 0.2, 0.4, and 0.6 s delayed.

Learning data for the NN are obtained from on-road tests including the curves to be analyzed with respect to accidents. Taking into account the nonlinear relationship between inputs x_i and output y for a single neuron with internal state u

$$u = \sum_i w_i x_i, \quad y = f(u), \quad f(u) = \frac{1}{1+\mathrm{e}^{-u}}, \tag{38.50}$$

the contribution of a respective input on the total output can be extracted. Proceeding with all neurons analogously, the amount and change of amount of input(s), for example, the information from

radar site (length and angle), to a particular output quantity, for example, the steering torque, with respect to time can be evaluated, for instance, with respect to its significance in the environmental recognition process.

The previously described method is extended to characterize the effects of experience and age on driving behavior in [110]. Further, a mental-workload model is proposed to be used to evaluate driver support systems.

The authors of [111,112] introduce a driver model that is established on a method for reinforced learning where the steering angle, as the control variable, is derived from the evaluation of a reward function. For several future time steps, states are estimated based on an internal vehicle model that is represented by a 3-layer NN. The reward depends on the (estimated) distance of the vehicle CG from the road borders.

A class of driver models that claim to mimic human driving behavior in particular are hybrid models (or integrated models). Although there is no common definition of this terminus, it is referred here to models that are composed by more than one modeling method to account for the complexity and multilayered human behavior, [12], on a psychological and physiological basis.

Early approaches are [113–115], the latter similar to the "STI compensatory driver model" [34], but considering thresholds to account for error tolerance behavior of the driver, see summary in [16]. Also [116] can be considered as an early attempt to describe and model different strategies of the driver.

At the hybrid driver model in [12], the previewed curvature ahead of the vehicle is preprocessed through NNs, resolving the curvatures into classes of "concepts" (straight line, curve, S-curve) [117], which are assembled to driver actions by FLC.

The hierarchical driver model presented in [118] is applied in 2-lane highway traffic simulations and is hybrid in the way that it divides the driver behavior into four layers, which are "decision making", "task planning", "maneuver", and "action". "Task planning", for example, for "overtaking", is typically rule based (if-then), while "action" is based on controllers, for example, preview controllers for "gap control" or "speed tracking". For avoiding any prompt danger, for example, a collision, a danger observer is set up. The execution of a "maneuver" will then be interrupted, the "decision-making" layer informed and, for example, the "avoidance" task selected. These safety judgment models are made up with the help of FLC.

A similar task of analyzing car-following and overtaking behavior of the human driver on a highway is performed in [119]. The hybrid model is composed of a "strategic", "tactical", and "operational" level, following Michon's description of driving behavior [9]. The "perception module" provides information on range and range rate with a leading vehicle (LV), which allows the simulated driver to choose a behavior among a set of possible behaviors, such as "following" or "overtaking". Part of the "tactical" module is the "driving behavior database", which allows for a quick retrieval of knowledge associated with a driving situation. In the "categorization" process, the current driving situation is matched with appropriate driving schema in the knowledge database; evaluation of aforementioned modules yields decision maps that define dangerous, critical, and comfortable regions as functions of range, range rate, and velocity. In the "operational" module, the feedback errors, for example, with respect to a desired time gap to the LV, are resulting from sliding mode controllers and are subsequently transformed to control inputs to the vehicle.

Such driver representations are called cognitive models. They embody both the abilities and constraints of the human operator, for instance, abilities such as memory storage and recall, learning, perception, and motor action and constraints such as memory decay, foveal versus peripheral visual encoding, and limited motor performance [120]. The architecture of cognitive models on the basis of an own model is described in more detail in [120]. Here, interesting in the context of the aforementioned controller-like "virtual test driver" models, a specific feature is mentioned. The control component of the driver model manages the perception of lower level visual cues and manipulates the vehicle's (longitudinal and lateral) controls. The lateral steering control is based on the perception of two visual points: The "near point" represents the vehicle's current lane position,

used to judge how close the vehicle is to the centerline of the roadway, and the "far point" indicates the curvature of the upcoming roadway used for anticipation and predictive compensation at a near-future position [121].

The hybrid and cognitive driver model in [122] is developed on the basis of theories on information processing [123,124]. Information processing, visual and vestibular, is described by two queueing systems. The visual information sources from traffic-relevant fixation points only generate an arrival event at the queue when the driver looks at a fixation point. An assumed relation, adapted to the different road fixation points, on the event rate at which the driver looks at the road fixation points, is used, typically 2–6 Hz. Events from fixation points inside the vehicle are modeled as disturbances. Each event is characterized by priority (urgency) and service time (how long the eyes look at fixation points without interruption). Due to queuing, events are sequentially served following priority; short-term memory is modeled with the strategy that all events that stay in the queue for longer than a specific period of time will be removed. The longitudinal and lateral controllers calculate their output (e.g., steering wheel angle) from the data in the short-term memory (e.g., velocity, lateral displacement), whether they are actual or not refreshed. The information for the vestibular perception queue is classified into longitudinal and lateral acceleration, yaw rate, and acceleration. Based upon the perceived driving situation, a suitable control action is selected from the driver's knowledge base (long-term memory). From there, reference values that depend on road layout and individual driver, for longitudinal and lateral control, are derived. The reference value (desired velocity) for the longitudinal controller, for instance, is determined by means of a finite state machine (Mealy automat) (Figure 38.13). Each state represents a different driving action. After the initialization at state "0", the state may switch to "5" if the vehicle is travelling in a curve (curvature $\kappa \neq 0$); otherwise, it switches into state "1", a straight path. The reader may continue intuitively in the graph (from "5" to "6", it is considered that the driver may accelerate before leaving the curve),

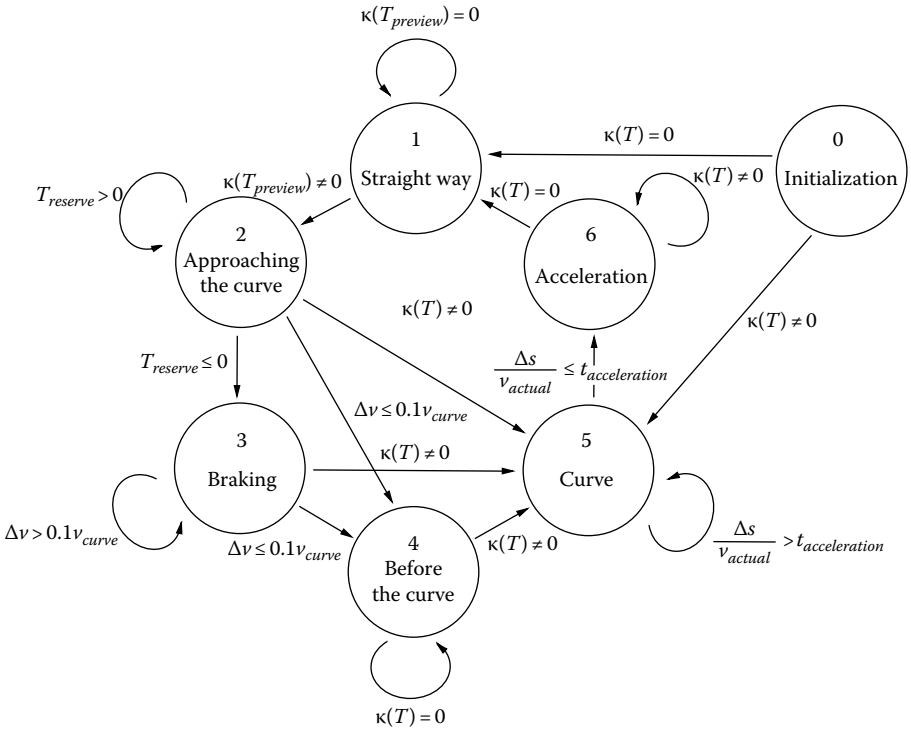


FIGURE 38.13 State graph of velocity controller. (From Kiencke, U. et al., *Control Eng. Pract.*, 7, 985, 1999; Kiencke, U. and Nielsen, L., *Automotive Control Systems*, Springer Verlag, Berlin, Germany, ISBN 3–540–23139–0.)

following his or her own driving experience (details in [122]). Finally, two discrete-time general predictive controllers (GPCs) are designed for longitudinal velocity control and lateral control to minimize the lateral displacement of the vehicle from a reference lane. The controllers calculate the suitable control variables that are accelerator pedal position, brake pedal force, and steering angle, from the information on the actual driving state, provided by the information acquisition and selection, the information on the future course path, and the reference values. More details on the driver model are to be found in [125].

For a further application based on the state-machine approach, see [126] and Section 38.3.2.

The change of driving strategies may also be represented by adapted identification methods. An example is the investigation of the stopping behavior of the driver in [127].

Authors in [128] propose that driver steering skill level is defined as the extent to which the driver possesses an understanding of the nonlinear vehicle dynamics. An internal model, which is considered to be a representation of the dynamics of the vehicle in the central nervous system, is used by the driver to predict the future vehicle motion within the preview horizon; the internal model concept is incorporated in different (mostly linear) ways in many of the driver models in Section 38.2. It is postulated in [128] that the driver learns and stores an understanding of the nonlinear vehicle dynamics by linearizing the dynamics around incremental operating points. The number of operating points corresponds then to skill and number of internal models used by the driver. For each linearized internal model, a predictive controller is designed and the most relevant selected and changed by the driver if applicable.

The behavior of drivers, especially in emergency situations, such as in the case of sudden encountering of an obstacle during driving, is analyzed in [129] for a better understanding of reasons of traffic accidents. The driver is mathematically described by a basic preview model in Laplace notation:

$$\frac{\delta_s(s)}{\Delta y(s)} = G_h(1+T_L s)\frac{1}{1+T_I s}, \quad \Delta y(s) = y_d(s) - \left[y(s) + vT_{pr}\psi(s) \right] \tag{38.51}$$

with the current and desired lateral displacement y and y_d, respectively. While parameters T_L and T_I are supposed to be constant, the steering gain G_h and preview time T_{pr} are considered to be variable under emergency situations. Their changes are obtained from searching for optimal solutions, based on the payoff function:

$$J = \int \left[a(y_d - y)^2 + b\delta_s^2 + c\dot{\delta}_s^2 + d\dot{G}_h^2 + e\dot{T}_{pr}^2 \right] \mathrm{d}t \tag{38.52}$$

with weighting factors *a, b, c, d, e*, by applying a genetic algorithm (GA). For avoiding the obstacle only by steering, a gene is composed of the distance between the vehicle and obstacle, the angle between the longitudinal axes of the vehicle, and the connection line between the vehicle and obstacle, the steering gain G_h, and the preview time T_{pr}; for the case of steering and braking, the gene is extended.

38.3.2 Path and Speed Planning, Optimized Driver/Driving Behavior

Path and speed planning as an internal driver task may be a priori incorporated in the driver model (see, e.g., [75] in this section or respective models in Section 38.3.1) or considered as a separate feature of the model. Many driver models (see Section 38.2) require a "desired path" to follow that is assumed to be given or to be the roadway centerline or to result just from a separate tool. For the latter case, some examples are cited here, whether they are part of a specific driver model or considered independently.

A desired path may be regarded as an optimized path. In this respect, either the total maneuver or course can be optimized (e.g., [68,70,130,131] and others as follows) or, from the driver point of view, toward an optimization horizon, some previewed seconds in advance [132,133], and after the driver has followed, optimization is started again. A variety of optimization criteria are treated in [132], which can be combined by respective weights to account for specific preferences by the driver:

Time optimal	To reach the destination as fast as possible
Acceleration optimal	To minimize horizontal accelerations, e.g., to enhance passenger riding comfort (to be combined with previous criterion to reach destination)
Brake optimal	To minimize the use of the brake to avoid unnecessary energy dissipation or brake disk overheating
Keep right/left	To stay in the preferred lane, if there is other traffic in the neighboring lanes
rpm optimal	To keep a certain engine speed, e.g., to avoid noise or to allow for enough engine power required for overtaking
Velocity optimal	To reduce the deviation from a given speed profile, e.g., to meet a posted speed limit

The optimization is performed considering basic prediction models of the vehicle's longitudinal and lateral dynamics and the tire behavior. A very similar approach is pointed out in [134], including considerations on the controller to track the optimized trajectory in [133].

The proposed path planning algorithm in [135] operates with polar curves applied to a line–arc–line transition at different speed conditions. A reference trajectory is derived that corresponds to the driver behavior, whose aim is assumed to avoid lateral acceleration variations during curve negotiation; therewith, a continuous curvature behavior constraint is guaranteed. The polynomial parameters of the reference trajectory are fuzzy (Mamdani based) determined with respect to database information (radius and turn angle of the curve) and specific constraints to account for different kinds of driving styles (novice, experienced driver) observed by driving tests. The path planning feature is a part of a navigation-aided driver-aid system, which locates the vehicle on the road by utilizing differential global positioning system (DGPS) and may either warn the driver, if an inappropriate driving condition in terms of the vehicle's speed and/or trajectory is detected, or may correct these situations.

The accuracy and generality of the driver trajectory model are improved in [136] by identifying the membership functions of the fuzzy interference system (now Sugeno based) from experimental data using an optimization algorithm.

The reference speed for the curve is determined considering the tolerated lateral acceleration (depending on the driver and vehicle) and on the instantaneous curvature of the trajectory followed and is thus not constant in the curve [126]. When negotiating the curve, it is assumed that the driver accomplishes a braking sequence to reach the reference speed for curve negotiation, a negotiation sequence nearly with a constant speed, and an acceleration sequence when leaving the curve depending on the road profile ahead. The cognitive mechanism is realized with a state machine, similar to Figure 38.13 (details in [126]).

In the context of speed choices in curves, the review [137] of various proposals thereof might be interesting as well (other proposals in [138]).

In [75], a driver model is constituted that incorporates the course decision process of the driver as well. It is assumed that the main information required for course decision is the risk feeling of the driver from a forward view. The risk level cannot be measured, but a high correlation thereof with the heart rate of the driver is presupposed. From experimental tests, it is concluded that the driver selects the course with minimum risk level. The risk level on a forward view is thus adopted and

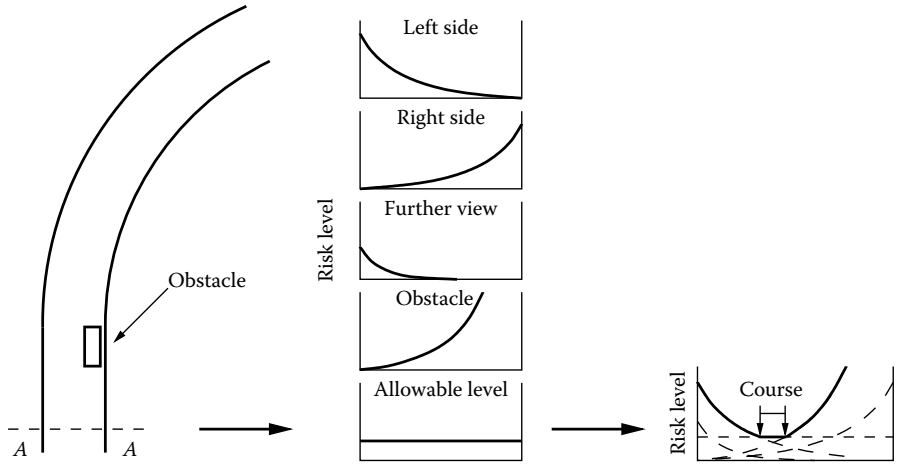


FIGURE 38.14 Driving course decision based on minimum risk level. (From Kageyama, I. and Pacejka, H.B., *Vehicle Syst. Dyn. Suppl*., 20, 314, 1991.)

composed of four parts that consider the distance from the left and right road border, obstacles, and the further view such as curves (Figure 38.14). These risk levels are described by exponential functions, whose parameters depend on velocity, boundary conditions, road width and road condition, etc., and are based on fuzzy reasoning.

The driver's steering angular velocity is then obtained from the risk level and its rate at a certain viewpoint ahead by applying FLC; from the actual steering angle and the desired steering angular velocity, a steering torque for the lateral control of the vehicle is derived. For the longitudinal control of the vehicle, respectively the decision of speed, the minimum risk level at the viewpoint is considered again.

To simulate a closed-loop ISO-lane change at the same maximum speed, an experienced human driver is able to perform, is not only a big, if not so far impossible challenge for a driver model. Finding a time-optimal trajectory, which the driver (model) has to follow (with time lag), is a solution to some extent but may fail on the extreme. The reason is the mismatch between the highly nonlinear vehicle behavior and an inadequate representation thereof by the driver model's internal prediction model and the fact that generally the driver behavior (e.g., time delay, limits) is ignored in the determination of the optimal trajectory. Considering the task in the latter way, different solution methods for the time-optimal maneuver or race circuit trajectory for the respective driver controls have been proposed.

In [139,140], parallel shooting method for solving the optimal control problem is applied. An optimized steering angle behavior is computed (by adaptive simulated annealing (ASA) method) in [141] for the ISO-lane change at a maximum speed performed by human test drivers; the same task is studied with a trajectory deformation method by iteratively deforming inputs in [142]. A generic algorithm is applied to the Customer Union (CU) short-course maneuver in [143]. The performance function in [144] for optimal steering and/or braking strategies at an obstacle avoidance maneuver consists of steering effort (steering angular velocity) and workload of the tires (combined friction coefficients) besides path constraints; direct shooting method is proposed for optimization.

Various performance indices with time-variant weighting coefficients on steering angle and angular velocity, yaw rate, roll rate, and deviation from road centerline, are adapted in [145] by trial and error to match the simulation results with the measured control action of a driver when passing through a double-lane change. The improved performance indices are then used as a quantification to explain the respective driving strategy of the driver.

In [146], the parameters of a vehicle and driver model, which is based on the observation of lateral displacement and yaw angle, are identified by applying an AR model for a closed-loop

single- and double-lane-change maneuver on experimental test data. In this way, an aimed separation of vehicle and driver behavior is realized.

The evaluation function for the optimization of a double-lane-change maneuver in [147] takes into account criteria for the minimum time to pass, the minimum margin between the border of the course and the corner of the vehicle, and the average steering wheel rate to represent the driver's physical effort; the predicted lateral displacement is used as input for an included basic driver model. Different weighting factors in the evaluation function are analyzed to conclude for intentions of the driver.

A driver model is developed for the ISO-lane-change maneuver in [148], where driving is seen as a model predictive control task in such a way that the driver accumulates knowledge about his or her vehicle's handling properties. Therefore, two (heuristic) fuzzy logic controllers are designed for the output variables steering wheel angular acceleration and accelerator pedal position. For lateral control, the input variables, position error, yaw error, and curvature error, are provided as a combination of their actual and predicted values and gains are determined to minimize the aforementioned errors.

In [149], also [150], two optimal driving policies are compared in particular at a limit handling maneuver, characterized by combined steering and acceleration, to achieve obstacle avoidance while maximizing speed and maintaining stability. The first, called general optimal control (GOC) technique, is working with vehicle states and road/obstacle geometry limits simultaneously, to provide the optimal sequence of steering, brake, and throttle actions to give the best obstacle avoidance. A nonlinear four-DOF handling vehicle model is employed, but there is no internal representation of the driver. The second, called dual-model (DM) technique, incorporates a generic driver model for closed-loop maneuvers and aims to provide near-optimal driver control inputs that are insensitive to driver model assumptions that may have an influence on the closed-loop performance evaluation. Therefore, a set of optimal controls is initially determined for a simple nonlinear particle model of the vehicle with mass-center acceleration constraints only. These results provide a reference input to a more realistic simulation that consists of a vehicle handling model coupled to a generic driver model:

$$\dot{u}_1 = \frac{(a_0 - \dot{u})}{\tau_1}, \quad \dot{u}_2 = k_1(\omega_0 - r) + k_2\sigma(\beta\dot{\beta})\dot{\beta},$$
$$\sigma(\beta\dot{\beta}) = \begin{cases} 1 & \text{if } \beta\dot{\beta} > 0 \\ 0 & \text{if } \beta\dot{\beta} \le 0, \end{cases} \tag{38.53}$$

with the acceleration command signal $u_1(t)$, which is converted to a drive force or braking torque, and the steering angle $u_2(t)$; u, r, β denote longitudinal velocity, yaw rate, and sideslip angle of the vehicle, respectively; τ_1, k_1, k_2 are the control parameters for demanded acceleration $a_0(t)$ and angular velocity $\omega_0(t)$ resulting from the particle model-based reference. From deviations between particle and driver–vehicle simulations, effective friction limits are estimated, which provide a modified set of conditions to revisit the reference optimization until no significant changes in controls are observed and an approximately optimal input is obtained.

38.4 APPLICATIONS WITH FOCUS ON VEHICLE/DRIVER COMBINATION

This section is concerned with application features that keep an eye on the combined vehicle/driver couple. The same tasks can be found in other sections as well, but while, for instance, the design of the controller for a certain driver support system may be the main concern for including a virtual test driver model into a simulation environment in Section 38.2, the application is presented in this section if reactions of these systems on the driver behavior or design criteria from the driver point of view, etc., are covered as well.

38.4.1 Vehicle Handling Dynamics with Respect to the Driver

The closed-loop system road–vehicle–driver is analyzed in [151] for an exemplary lane-change maneuver. The capabilities of the driver, as a basic controller of the previewed positional error, are limited by maximum attainable steering wheel angular movement parameters. If any of these limits are violated, the model continues in an open-loop mode, where the steering wheel movement is directed by the maximum possible movement parameters. Emergency trajectories are generated until the driver steers again in closed-loop mode or crashes. For the presented maneuver, evaluation criteria are stated both with respect to performance and to the driver's

Mental effort:

- Based on the abruptness of vehicle movements, like maximum yaw angles and yaw velocities, and sideslip angles of the vehicle
- Based on the task-dependent driver's foresight (phase angle difference between steering wheel motion and individual vehicle reaction), like width of the hysteresis loops of $\dot{\psi}/\delta_s$, $\dot{\psi}/a_y$, β/a_y, and integral phase lag criterion computed as integral of the difference between normalized steering angle and normalized curvature of the CG's trajectory, etc.

Physical efforts:

- Like maximum steering wheel velocities, accelerations and steering wheel torque, total angular movement of the steering wheel, mean steering wheel velocity, maximum steering wheel power, mean positive steering wheel power

In [152], an analytical approach to predict the subjective rating of handling qualities of actively controlled vehicles is presented and experimentally tested for the evaluation of 2/4-wheel-steered vehicles. Handling qualities are understood here as qualities of the vehicle that affect ease and accuracy in performing assigned driving tasks and imply thus the driver, in particular task performance and driver's workload. The transfer function–based driver model

$$\delta(s) = K_{\psi}(T_{L\psi}s+1)\left[K_y(T_{Ly}s+1)y_e(s) - \psi(s)\right]\frac{e^{-\tau_r s}}{(T_I s+1)} \tag{38.54}$$

with the lateral position error y_e and yaw angle ψ includes the gains K_y, K_{ψ} and the lead time constants $T_{Ly}, T_{L\psi}$ to account for the adaptation of the driver to the vehicle characteristics and assigned task, while the time delays τ_r, T_I are assumed not to be subjected to any changes due to the driver's intention. The driver's subjective rating is represented by an objective function J with appropriate weights q_1, q_2, q_3:

$$J = q_1\int_0^{t_f} y_e^2 dt + q_2\int_0^{t_f}\delta^2 dt + q_3(T_{Ly}+T_{L\psi}) \tag{38.55}$$

that takes into account the task performance (accuracy) and the driver's physical (steering effort) and mental workload (required lead time constants). The parameters K_y, K_{ψ}, T_{Ly}, $T_{L\psi}$ that minimize the value of J are determined by numerical optimization, and J_{min} serves as a predicted value for the subjective ratings in accordance with experimental tests. Specific design parameters that influence vehicle handling behavior are then evaluated with respect to their effect on the rating J_{min} and its constitutive elements.

By analyzing driver model and vehicle behavior, a design strategy for a steer-by-wire (SBW) system is presented in [153] to improve vehicle stability while preserving handling performance during

lane change at a given constant speed. Therefore, a driver model, which incorporates both steering wheel angle δ_s and steering torque M_d, is introduced by

$$\frac{\delta_s}{K_\delta} + \frac{M_d}{K_M} = \Delta y_{pr} e^{-\tau_r s} \quad \text{and} \quad \Delta y_{pr} = y_d - (y + L\psi) \tag{38.56}$$

with future lateral position error Δy_{pr}, desired lateral position y_d, driver's delay time τ_r, preview distance L, current lateral position y, and yaw angle ψ. For $K_M = \infty$, the driver controls the vehicle with the steering angle only; for $K_\delta = \infty$, with the steering torque only; and for $0 < (K_\delta, K_M) < \infty$, with a combination of these. It is further assumed that $e^{-\tau_r s} \approx 1/(1 + \tau_r s)$ and that moment of inertia and viscous friction of the steering system are small and negligible; thus, $M_d = -M$, where M is the steering torque from the vehicle. The addressed improvements derive from an optimized steering angle gain k (with a performance function based on lateral position error and lateral acceleration of the vehicle) and the compensation of the system delay time, made up by vehicle and driver delay. To attain this objective, the free design parameters of the SBW system are chosen to realize lead steering, $\delta_f(s) = k(1 + \tau_{rs})\delta_s(s)$ (with front wheel steering angle δ_f), and/or feedback of the driver's steering torque derivative including the gains K_δ, K_M and the time delay τ_r of the driver model in (38.56).

In [154], a driver-articulated vehicle model is described to study and adapt the vehicle design with respect to different driver skills. Figure 38.15 illustrates the proposed multiloop structure of the driver/vehicle system, where the path input may be either the lateral displacement or the path orientation.

The inner control loop signals comprise the lateral acceleration of the lead unit, articulation rate of the combination, and roll angle of the sprung mass of the trailing unit to employ sensitivity of the driver to both tractor and trailer. The components of the compensatory function of the driver are $H_L(s) = K_L e^{-\tau_L s}$ and $H_H(s) = K_H e^{-\tau_H s}$, respectively; the preview block $P(s) = e^{T_p s}$ represents the preview function for a single-point preview strategy. The driver's prediction function for the lateral displacement $B(s)$, like stated in (38.2) before, represents the driver's prediction ability of the vehicle motion. The prediction functions in the internal loops are defined by $F_i(s) = K_{pi} e^{\tau_{pi} s}$ ($i = 1, 2, 3$) and are intended to indicate the driver's direct and/or indirect perception of the vehicle motion: For direct perception of the motion arising from either experience or subconscious knowledge of vehicle response, τ_{pi} assumes a positive value; indirect perception of motion, realized through monitoring the vehicle state, yields a negative value of τ_{pi}, since the delays associated with measurement, signal processing, display, and detection are longer than the driver's prediction time. It is further postulated

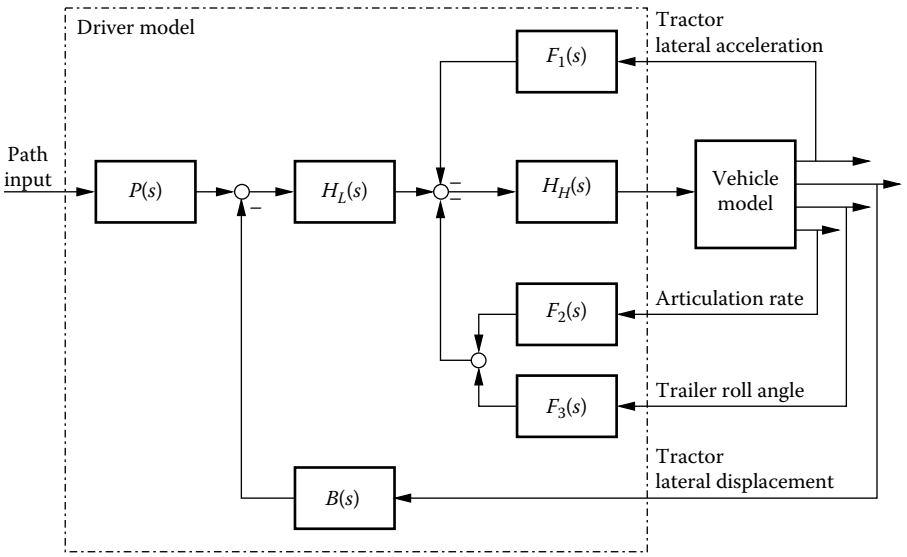


FIGURE 38.15 Driver model for articulated vehicles. (From Yang, X. et al., *J. Mech. Des.*, 123(1), 132, 2001.)

that the driver's control limits in terms of reaction time, preview distance, compensatory gain, and steering input vary considerably with individual skills, test conditions, and environmental factors.

A performance index based upon the vehicle path tracking, the directional response characteristics, and the driver's steering effort is formulated and minimized to derive the desirable ranges of vehicle parameters that could be adapted for drivers with varying skills. A detailed description of driver model for articulated vehicles is given in [155].

38.4.2 Accident Avoidance, Active Safety, Driver Support Systems

Similar to the just previously referenced task in [153], the adaptation of SBW parameters to the steering characteristics of (in particular) aged drivers is investigated in [156]. For that purpose, the parameters of the driver model according to (38.51), but with $\Delta y(s) = y_d(s) - y(s)$, are determined by minimizing a cost function composed of the deviation between calculated and measured data for lane-change maneuvers; the target path y_d is considered as a step input. A further optimization (for minimal deviation between actual and desired lateral distance) is then applied, to find the parameters of the feedforward SBW system for the closed-loop system to follow an ideal vehicle trajectory. For the lane-change maneuver, a double integration of a pure sinusoidal acceleration, with the magnitude corresponding to the lane width and the frequency corresponding to lane-change distance and velocity, is chosen as an ideal path.

A driver model based on multiple inverse models of the dynamics of the vehicle is used in [157] for the design of a driver support system to assist the driver in an obstacle avoidance scenario. A general description of an "integrated driver model" for driver safety assist systems is discussed in [158].

The focus of the study [159] is on the adaptation process of driving behavior when using an adaptive cruise control (ACC) system. ACCs are capable of automatically maintaining a constant intervehicle distance and speed, which are selected and set by the driver. To account for different individual extents of skilfulness of drivers, a feedback-error learning scheme is introduced as driver model based on an NN and two feedback proportional controllers. The inputs of the NN are the road curvature, a preview distance ahead; the velocity of the vehicle; the difference between desired and actual distance to the lead vehicle; steering angle; lateral deviations from the road centerline (at three distances ahead); and the on–off signal for ACC use; output quantities are steering angle and velocity of the vehicle. The simulation model is then used for predicting the control performance of a skilled driver using ACC. Headway time for lane changing in a row, FFT analysis of the steering angle, and lateral deviation from the road centerline are considered for driver behavior description during ACC use and manual driving, respectively.

The results of the development of a driver assisting system as a monitoring and warning aid for the driver on highways are presented in [160–162]. The basic outline of the system is given in [161]. The "reference driver module" (based on state automata in [161], on petri nets in [162] representing a statistically average driver behavior) extracts from the incoming data of machine vision and sensors a symbolic description of the actual situation. In addition, it calculates and outputs situation-specific action boundaries. The "discrepancy interpretation module" judges the driver's actions with respect to these boundaries, thereby also taking into account the situation-specific output of the "model of actual driver." In case of an actual danger coming up, the "warning device" is triggered to issue an appropriate warning, which is presented to the driver through haptic displays. The "model of actual driver" accounts for the adaptivity to the individual driver, where the actual driver's normal driving style is extracted from observations of his or her behavior during a learning period. Therefore, statistical approach and NN approach are compared in [160,161], and an NN approach based on fuzzy adaptive resonance theory (ART) architectures is introduced in [162].

In [163], "hidden Markov dynamic models" (a set of dynamics models sequenced together by a Markov chain) are used to recognize human driving behavior from sensory data and to predict the most likely driver's behavior over three seconds; the human is considered as a device with a large number of internal mental states, each with its own particular control behavior and interstate

transition probabilities. By observing the temporal pattern of the driver's control action behavior and comparing it to the action models, the initiated action of the driver can be determined and may be used for collision avoidance systems ([164], also [165,166]). A further approach for predictive action recognition applying probabilistic networks for ACC is outlined in [167]. A model of driver intention recognition based on a fuzzy associative memory system is described in [168].

A kind of hybrid driver model is proposed in [129,169] to simulate and understand the driver's strategy at the instance of a sudden stopping of the preceding vehicle. A set of linear polynomials is used to describe the steering amount based on sensory information of range and range rate of the vehicles and their lateral displacements. The coefficients of these polynomials as well as the parameters, which define the switching condition between the "modes" with respect to the range between the vehicles, are identified from data and observed stopping strategies using a driving simulator environment. Driving characteristics of the novice and expert driver are compared.

The use of steering control as a driver assistance function in collision situations is addressed in [170]. For that purpose, a basic preview steering control is applied to stabilize the trajectory of the vehicle at postimpact maneuvers.

In [171], the gear selection problem of automatic transmissions (ATs) is analyzed, and a learning algorithm is proposed to improve the vehicle behavior and driver satisfaction. To optimize vehicle performance with multiple antagonistic criteria, locally and globally over time, a parametric discriminant function is considered, depending on an evaluation of the driver satisfaction, and the so-called driver-style state functions, as a reward for the system, and a reinforcement learning algorithm are applied, derived from Q-learning method. Supervised learning directly from the driver is performed when the driver selects AT ratio in manual mode; that allows the embedded controller to customize memorized strategies when resuming AT in automatic mode (based on a backpropagation NN). Each driver may possess a personal driving style and may also be characterized by different physical states, which are represented by a class creation/selection mechanism for unsupervised learning.

38.5 APPLICATIONS WITH FOCUS ON THE ENVIRONMENT/TRAFFIC

The final section is concerned with applications of driver models that consider the environment of the vehicle–driver combination, in particular the interaction with one or more other vehicles. In this respect, topics such as understanding the system "traffic", traffic flow, and car following come into view. Car following itself forms one of the main processes in all microscopic simulation models as well as in modern traffic flow theory, which attempts to understand the interplay between phenomena at the individual driver level and global behavior on a more macroscopic scale [172]. Applications like ACC and automated highway systems (AHSs) entailed a rich series of publications and as long as the "driver" is the main concern and not automated driving (or "intelligent cruise control" that may consider only partial aspects of the human driver); some early elementary as well as some recent papers are presented in the following. Giving an overview of some principal modeling methods and typical applications, rather than insight, is attempted here; related models and topics have also already been addressed before.

Already in 1950, the longitudinal behavior of a traffic line or platoon of vehicles was investigated by Reuschel [173] on a kinematic basis. The LV initiates the longitudinal states of the following vehicles (FVs) at a given rule for the gap Δs between the vehicles:

$$\Delta s(t) = \Delta s_0 + qv(t) \tag{38.57}$$

with a predefined gap Δs_0 between preceding and FV at standstill and velocity $v(t)$ of the FV; q is fixed to 3.6 (for the conversion of v in km/h to m). The case of a sudden deceleration to standstill and the effect on the following platoon has been treated as well. Since then, the topic attracted a growing interest along with growing traffic, and more complex models to simulate traffic flow have been developed.

A longitudinal driver model for simulating traffic lines is included in [174]. Considering transfer functions for the longitudinal vehicle dynamics and driver behavior

$$\frac{s\alpha(s)}{\Delta v(s)} = \frac{k_M}{s}\frac{(1+T_{L,1}s)(1+T_{L,2}s)}{1+T_N s} \tag{38.58}$$

with the accelerator pedal position $\alpha(s)$, range rate $\Delta v(s)$, and driver parameters k_M, $T_{L,1}$, $T_{L,2}$, T_N, stability aspects of traffic lines are analyzed.

The simplified version of the "precision model" (see (38.7) [30]) is proposed in [175] to model driver behavior for noncritical car-following situations. The identification of the driver parameters is presented as well as evaluation with experimental data.

The determination of the range/gap between two successive vehicles as a kind of rule or demand characteristic for the driver (and often referred to as longitudinal "driver model") in car-following situations has been studied theoretically and experimentally in numerous publications. In 1953, a linear transfer behavior is suggested in [176] by specifying the acceleration $\dot{v}_{k+1}$ of the vehicle $(k + 1)$ proportional to the relative velocity $(v_k - v_{k+1})$ to the preceding vehicle k. The idea is extended in [177] by a time lag τ_r (reaction time of the driver), so the basic leader–follower model reads

$$\dot{v}_{k+1}(t) = \lambda\left[v_k(t-\tau_r) - v_{k+1}(t-\tau_r)\right] \tag{38.59}$$

with constant λ called the sensitivity factor of the driver. One example of the numerous applications of the model is [178], where the effects on traffic flow characteristics (string stability) and environment (fuel consumption) are analyzed, when semiautomated vehicles with automatic vehicle-following capability operate together with manually operated vehicles.

The nonlinear type of this model became well known and is often referred in literature as GM model or Gazis–Herman–Rothery (GHR) model [179]:

$$\dot{v}_{k+1}(t) = C v_{k+1}^m \frac{v_k(t-\tau_r) - v_{k+1}(t-\tau_r)}{\left[x_k(t-\tau_r) - x_{k+1}(t-\tau_r)\right]^l} \tag{38.60}$$

For suggestions on the choice of the parameters C, l, m, including some references of many published modifications of the model, see [172] or more recently [180].

Another modification, known as (simplified) Helly model [181]

$$\begin{aligned}\dot{v}_{k+1}(t) = {} & C_1\left[v_k(t-\tau_r) - v_{k+1}(t-\tau_r)\right] \\ & + C_2\left[(x_k(t-\tau_r) - x_{k+1}(t-\tau_r)) - D(t)\right]\end{aligned} \tag{38.61}$$

$$D(t) = \alpha + \beta v_{k+1}(t-\tau_r) + \gamma \dot{v}_{k+1}(t-\tau_r) \tag{38.62}$$

with a desired following distance $D(t)$ and where C_1, C_2, α, β, γ are parameters to specify the characteristic of the traffic flow, is also used up to now, for example, even in a neural fuzzy framework in [182].

A state-space formulation is chosen in [183] to derive with optimal control theory constant feedback gains, which modify the dynamics of the follower in a way that the resultant system response approaches that of a given model (of the leader). Examples are presented for aircraft dynamics. The method is transferred to the car-following problem in [184]. A quadratic performance criterion function guarantees that

deviations from some desired spacing $\Delta s_d = kv_f$ and the relative velocity Δv are minimized and that minimum energy is required; then the optimal control becomes (with the speed v_f of the follower)

$$u(t) = C_s[\Delta s(t) - C_c v_f(t)] - C_v \Delta v(t) \tag{38.63}$$

After including the driver's reaction time τ_r and vehicle nonlinearities (which are not specified any further), the parameter set C_s, C_c, C_v, τ_r is identified from highway data. With a switching method also, more preceding vehicles may be considered.

Another often cited car-following model to mimic real traffic was presented by Gipps in 1981 [185]. When it is assumed that the driver travels as fast as safety and limitations of the vehicle permit, the speed v_n of vehicle n at position x_n is given at time $(t + \tau_r)$ by

$$v_n(t+\tau_r) = \min \begin{cases} v_n(t) + 2.5a_n\tau_r\left(1 - \dfrac{v_n(t)}{v_{n,0}}\right)\sqrt{0.025 + \dfrac{v_n(t)}{v_{n,0}}} \\ b_n\tau_r + \sqrt{b_n^2\tau_r^2 - b_n\left[2(x_{n-1}(t) - s_{n-1} - x_n(t)) - v_n(t)\tau_r - \dfrac{v_{n-1}(t)^2}{b}\right]} \end{cases} \tag{38.64}$$

The parameters required for the model are listed in Table 38.1.

The first constraint to vehicle n in (38.64) is that it will not exceed its driver's desired speed and its acceleration should first increase with speed as engine torque increases and then decreases to zero as the vehicle approaches the desired free flow speed $v_{n,0}$.

The second constraint takes into account that the driver of the FV selects his or her speed to ensure that he or she can bring his or her vehicle to a safe stop should the vehicle ahead came to a sudden stop. That's the way models of this type are called safety distance or collision avoidance models; an early contribution of this type is [186].

In this way, the model combines the two competing goals of keeping a safe distance from the preceding vehicle and to converge to the driver's desired free flow speed.

The approach in [187] is based on (38.59) ($\tau_r = 0$) where v_k is replaced by an optimal velocity function $v_{opt}(\Delta x)$:

$$v_{opt}(\Delta x) = v_0\left[\tanh m(\Delta x - b_f) - \tanh m(b_c - b_f)\right] \tag{38.65}$$

with headway $\Delta x = x_{k+1} - x_k$ and the parameters v_0, m, b_f, b_c to be determined from observed data of the individual behavior of vehicle k.

TABLE 38.1
Parameters of Gipps Car-Following Model

a_n	Max. acceleration that the driver of vehicle n wishes to undertake
b_n	Max. deceleration that the driver of vehicle n wishes to undertake
$\hat{b}$	Estimated value of b_{n-1}
$v_{n,0}$	Speed at which the driver of vehicle n wishes to travel
s_n	Save distance to vehicle $n-1$
τ_r	Reaction time

Source: Gipps, P.G., *Transport. Res. B*, 15B, 105, 1981.

Validation of the aforementioned models [177,179,184,185,187] as well as the car-following model for traffic flow in [188] is done on the basis of two databases with field measured car-following data in [189]. The Gipps model is reported to be able to fit the highest number of maneuvers; nonetheless, in [55], an improved version of the Gipps model has been proposed. In particular, when traffic density is high, and the vehicle behavior is dominated by the second constraint in (38.64), the Gipps model is corrected not to overestimate traffic flow. The updated model is then used to evaluate the impact of ACC vehicles on traffic flow (see also [190]).

An overview of a similar group of car-follower models as mentioned earlier is discussed in [191] with respect to stability issues. The authors criticize on missing human factors in the models and refer to [192], where car following is described from a physiological/psychological point of view; a model is presented in [193]; [192] is written as a comment to [172] that gives a historical review on car-following models from the beginnings to 1999. In return, [194] is a reply on [192] and focuses on the implications of satisficing, the determinism assumed in engineering-inspired models of car following, and the role that individual differences in perceptual and cognitive performance, as well as motivation, must play in future accounts of traffic behaviors such as car following.

Other comparative studies including previous or similar type of models for traffic flow or ACC investigations are, for example, [195,196] or [197].

Traffic simulation models make assumptions about the safety-related behavior of drivers. The authors of [198] state that these assumptions may or may not replicate the real behavior of those drivers, who adopt seemingly unsafe behavior, for instance, when following the vehicle in front too closely. Consequently, the paper explores the questions associated with the choice of values for safety-related parameters of various simulation models in literature.

The transition from free traffic flow to synchronized flow at a highway bottleneck is studied with two empirical microscopic traffic flow models in [199].

With respect to other or more specific applications, [200] is noted, where the driver of a heavy-duty vehicle is considered following another vehicle. Based on measured data, a basic feedforward and feedback model is developed. The applied multiple linear regression method (similar to the driver model for steering in [110]) is used to set a linear relationship between input variables (information on preceding vehicle, FV, and relative to each other) and an output (throttle position); derived coefficients are then analyzed with respect to the importance of each information to the driver.

The behavior of non-truck drivers is modeled in terms of their interactions with trucks in a traffic stream in [201]. The car-following and lane-changing logics of the microscopic highway traffic flow model [202] are therefore extended. Based on the survey data, the car–truck interactions are described with a fuzzy logic–based model by associating a "discomfort level" for the non-truck driver in the vicinity of trucks.

In [203], night driving behavior is investigated. For the optimal velocity v_{opt} in the car-following model

$$\dot{v}_{k+1} = C_a[v_{opt}(\Delta x) - v_{k+1}] + C_b[v_k - v_{k+1}] \tag{38.66}$$

with parameters C_a, C_b, the same approach as given in (38.65) is chosen but adapted with thresholds. Therewith, stability analysis of traffic clusters is done.

The conceptual paper in [204] shows another kind of application for driver models. A driving simulator and a macroscopic traffic simulator are linked together in real time, interacting with a further microscopic traffic simulator. Therein, a driver model is included to generate response to traffic environment with other vehicles, respectively, to interfere with the real driver in the driving simulator.

Longitudinal car-following driver models have also been introduced in driver warning systems for vehicle collision avoidance (e.g., [205]) similar to the objectives for the lateral case in Section 38.4.2.

A different kind of modeling approach for the car-follower problem is the application of NNs [108], which has already been outlined in Section 38.3.1. The recent study in [206] compares the efficiency of a backpropagation NN and a radial basis function (RBF) NN for modeling car-following behavior. Also FLC has been used to develop knowledge-based models of traffic behavior (e.g., [201,207]).

In [208], a stochastic modeling of mixed traffic flows on a multilane road is presented. For saturated flows, in a "stop and go" mode, it is taken into account that a great number of vehicles may change to another lane. Stability aspects are discussed, and methods to estimate energy consumption and intensity of accompanying pollutant emissions are proposed.

In constitutive equations of basic car-following models (e.g. (38.60)), it is presupposed that the driver of the following car reacts, on the one hand, to arbitrarily small changes in the relative speed and, on the other hand, even at very large spacings [209]. Therefore, in these equations, it is assumed that there is no response as soon as speed differences disappear, which is considered not realistic in [209], and "psychophysical" models are proposed there instead: Drivers are subject to certain limits on the stimuli to which they respond; for example, at large spacings, the driver of an FV is not influenced by the amount of the speed difference. Only when these perceptual thresholds are reached, the driver will be able to perceive the change in the apparent size of the LV and react to changes in the kinematic variables; the achievements presented in [210] are based on this idea, also in [211]. At close headways, where the speed difference is likely to be below a threshold, spacing-based thresholds become relevant ("action point" models).

An (if-then) rule-based cognitive model is realized in [212] to give an example for a descriptive modeling approach.

Taking into account that drivers rely chiefly upon visual information, the visual system's ability to detect the onset of the LV's braking and to monitor and track the FV's approach to the LV is studied and quantified in [213]; an early contribution in this respect is [214]; see also [215] and references in [213]. Visual cues include

- Monocular expansion of the LV's image in the FV driver's field of view (looming)
- Texture dilation (as the LV approaches the FV, the fine detail of its surface features comes more sharply into focus)
- Binocular cues
- LV brake light ignition

In [213], the looming cue is focused exclusively, represented by the rate of change of the angle θ as defined in the right plot of Figure 38.16.

The human driver model of the LV braking scenario is illustrated in Figure 38.16. The sensory input to the visual system, the monocular image expansion, is represented by $\dot{\theta}$ with added noise. Pure time delays are also associated with the visual system that is provided by the block $e^{-s\tau}$. Then the signal passes through a filter. The FV driver's perception of $\dot{\theta}, \dot{\theta}_p$ is then sent to the "cognitive process" block, along with a judgment by the "criterion detector" as to whether LV braking is present. If $\dot{\theta}_p$ exceeds C, the "criterion detector" judges that LV braking has been initiated and a braking strategy is initiated. Otherwise, some other strategy (e.g., car following) is pursued. The criterion value used in assessing $\dot{\theta}_p$ is a function of the expected costs and benefits of the decision outcomes, which in general will depend upon the particular driving conditions, the observation noise, and the minimum expansion rate to be detected. Neuromuscular models of human leg and foot movement will be incorporated into the cognitive and neuromuscular blocks in Figure 38.16, completing the human driver model.

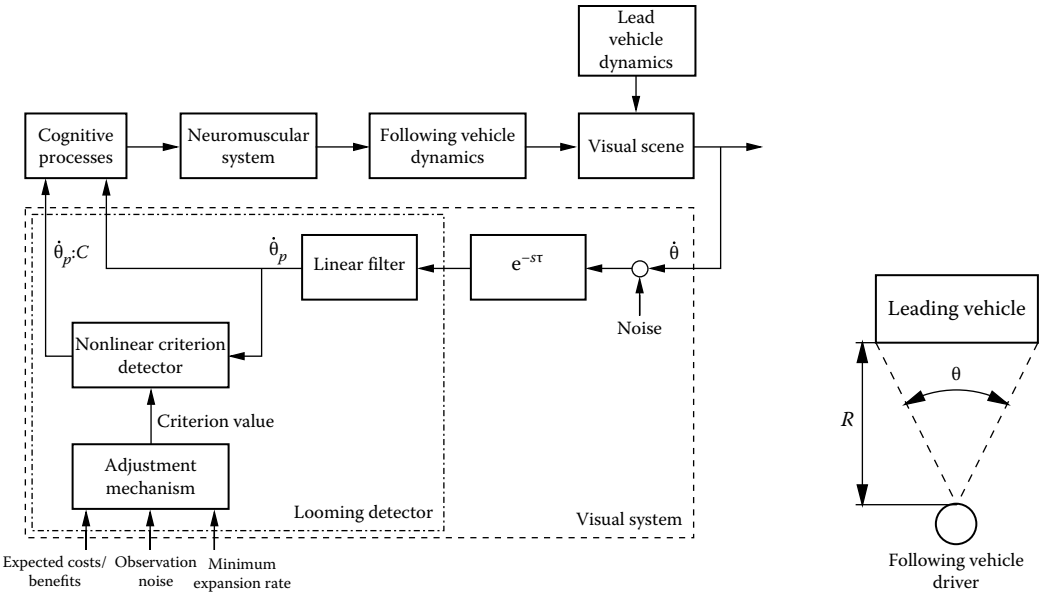


FIGURE 38.16 Human driver model of LV braking scenario and definition of visual angle θ. (From Barton, J.E. et al., Towards a complete human driver model: The effect of vision on driving performance, *Proceedings of the American Control Conference*, Minneapolis, MN, pp. 2591–2598, 2006.)

38.6 CONCLUSION

The reviewed articles reveal a very wide spectrum of mathematical methods applied to model the driver for understanding and improving automobile dynamics and for studying human driving behavior. Technical issues are predominant with respect to application, and it may be noted that the human factor in the coupled system vehicle–driver receives not always adequate attention. In a way, the technical understanding of the automobile is still far beyond the representation of the *human* driver. Nevertheless, physiological and psychological considerations determine many driver models to a certain extent, and much effort can be noticed in incorporating knowledge from human sciences to compensate for the mathematical theories and more technical understanding of driving. Although experimental validation lags somewhat behind the advanced mathematical methods, driver models became considerably important in VSD.

The wide variety of mathematical methods applied to determine the behavior of the driver may not only result from particular preferences of the individual research scientist but follow the wide variety of different areas of application, as could be observed in this chapter. Conclusions from an application to a conclusive mathematical method may not be drawn, although some specific modeling approaches seem to be more effective with respect to the number of applications and publications than others, but *the* driver model for a certain kind of application is not apparent as it is the case for other models in VSD. And for good reason, driving is a complex physical and mental individual process and will keep scientists interested and inventive to simulate and understand human (driving) behavior.

REFERENCES

1. Riekert, P. and Schunk, T., 1940, Zur Fahrmechanik des gummibereiften Kraftfahrzeugs. *Ingenieur-Archiv*, **11**(3), 210–224, (in German).
2. Kondo, M., 1953, Directional stability (when steering is added). *Journal of the Society of Automotive Engineers of Japan (JSAE)*, **7**(5/6), pp. 104–106, 109, 123, 136–140 (in Japanese).

3. Iguchi, M. et al., 1960, A fundamental study on characteristics of driver control action with a model automobile. *Journal of the Society of Automotive Engineers of Japan (JSAE)*, **7**, (in Japanese).
4. Sheridan, T.B., 1963, Vehicle handling: Mathematical characteristics of the driver. *Society of Automotive Engineers (SAE)*, **630068**, 1–6.
5. Enke, K., 1966, Überlegungen zum Zusammenwirken von Fahrzeuglenker und Automobil. *Automobil-Industrie*, **1**, 37–44, (in German).
6. Fiala, E., 1966, Lenken von Kraftfahrzeugen als kybernetische Aufgaben. *Automobiltechnische Zeitschrift (ATZ)*, **68**(5), 156–162, (in German).
7. Wierwille, W.W., Gagné, G.A., and Knight, J.R., 1967, An experimental study of human operator models and closed-loop analysis methods for high-speed automobile driving. *IEEE Transactions on Human Factors in Electronics*, **HFE-8**(3), 187–201.
8. Rashevsky, N., 1966, *Neglected Factors in Highway Safety*. University of Michigan Mental Health Research Institute, Ann Arbor, MI, Grant GM-12032–01.
9. Michon, J.A., 1985, A critical review of driver behavior models: What do we know, what should we do? In: R. Schwing and L. A. Evans (Eds.), *Human Behavior and Traffic Safety* (pp. 487–525) (New York: Plenum Press).
10. Miall, R.C. and Wolpert, D.M., 1996, Forward models for physiological motor control. *Neural Networks*, **9**(8), 1265–1279.
11. Rix, J.J. and Cole, D.J., 2002, Models of human learning applicable to the vehicle steering task. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2002*, Hiroshima, Japan, pp. 683–688.
12. Jürgensohn, T., 1997, Hybride Fahrermodelle. (Berlin, Germany: Pro Universitate Verlag), (ISBN 3–932490–22–3), (in German).
13. MacAdam, C.C., 2003, Understanding and modeling the human driver. *Vehicle System Dynamics*, **40**(1–3), 101–134.
14. Willumeit, H.-P. and Jürgensohn, T., 1999, Fahrermodelle—ein kritischer Überblick, Teil 1 und 2. *Automobiltechnische Zeitschrift (ATZ)*, **99**(7/8, 9), 424–428, 552–560, (in German).
15. Hoffmann, E.R., 1975/1976, Human control of road vehicles. *Vehicle System Dynamics*, **5**, 105–126.
16. Reid, L.D., 1983, A survey of recent driver steering behavior models suited to accident studies. *Accident Analysis and Prevention*, **15**(1), 23–40.
17. Irmscher, M., Jürgensohn, T., and Willumeit, H.-P., 1999, Driver models in vehicle development. *Vehicle System Dynamics Supplement*, **33**, 83–93.
18. Riedel, A., 1990, IPG-Driver—Ein Modell des realen Fahrers für den Einsatz in Fahrdynamik-Simulationsmodellen. *Automobilindustrie*, **6**, 655–662, (in German).
19. Guo, K. and Guan, H., 1993, Modelling of driver/vehicle directional control system. *Vehicle System Dynamics*, **22**, 141–184.
20. Kondo, M. and Ajimine, A., 1968, *Driver's Sight Point and Dynamics of the Driver-Vehicle-System Related to It*. Society of Automotive Engineers (SAE), New York, 680104.
21. Jürgensohn, T., 2006, Control theory models of driver behaviour. *Proceedings of the Workshop Modelling Driver Behaviour in Automotive Environments*, May 2005, Ispra, Italy, pp. 25–27.
22. Chatelet, P. and Anhtuan, P., 1973, Dynamique du système conducteur-véhicule. *Ingenieurs de l'automobile*, 757–812, (in French).
23. Leglouis, T., Laneville, P., Bourassa, P., and Payre, G., 1986, Characterization of dynamic vehicle stability using two models of the human pilot behaviour. *Vehicle System Dynamics*, **15**, 1–18.
24. Leglouis, T., Laneville, P., Bourassa, P., and Payre, G., 1987, Vehicle/Pilot system analysis: A new approach using optimal control with delay. *Vehicle System Dynamics*, **16**, 279–295.
25. Yoshimoto, K., 1968, Simulation of driver/vehicle system including preview control. *Journal of Mechanics Society Japan*, **71**, (in Japanese).
26. Mitschke, M., 1993, Driver-vehicle-lateraldynamics under regular driving conditions. *Vehicle System Dynamics*, **22**, 483–492.
27. Appel, A. and Mitschke, M., 1997, Adjusting vehicle characteristics by means of driver models. *International Journal of Vehicle Design*, **18**(6), 583–596.
28. Reichelt, W., 1990, Ein adaptives Fahrermodell zur Bewertung der Fahrdynamik von Pkw in kritischen Situationen. PhD thesis, Braunschweig University of Technology, Braunschweig, Germany, (in German).
29. McRuer, D.T. and Krendel, E.S., 1959, The human operator as a servo system element. *Journal of the Franklin Institute*, **267**(5/6), 381–403, 511–536.
30. McRuer, D.T., Graham, D., Krendel, E.S., and Reisener, W., 1965, Human pilot dynamics in compensatory systems: Theory, models and experiments with controlled-element and forcing function variations. Wright-Patterson Air Force Base (OH): Technical Report AFFDL-TR-65–15.

31. Weir, D.H. and McRuer, D.T., 1970, Dynamics of driver vehicle steering control. *Automatica*, **6**, 87–98.
32. McRuer, D.T. and Krendel, E.S., 1974, Mathematical Models of Human Pilot Behavior. Systems Technology, Inc., Hawthorne, CA, AGARD AG 188, STI-P-146.
33. McRuer, D.T., Allen, R.W., Weir, D.H., and Klein, R.H., 1977, New results in driver steering control models. *Human Factors*, **19**(4), 381–397.
34. McRuer, D.T., Weir, D.H., Jex, H.R., Magdaleno, R.E., and Allen, R.W., 1975, Measurement of driver-vehicle multiloop response properties with a single disturbance input. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-5**(5), 490–497.
35. Reid, L.D., Graf, W.O., and Billing, A.M., 1980, The validation of a linear driver model. University of Toronto Institute for Aerospace Studies (UTIAS) Rept. No. 245.
36. Allen, R.W., Szostak, H.T., and Rosenthal, Th.J., 1987, Analysis and computer simulation of driver/vehicle interaction. *Society of Automotive Engineers (SAE)*, 871086.
37. Lin, M., Popov, A.A., and McWilliam, S., 2002, Handling studies of driver-vehicle systems. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2002*, Hiroshima, Japan, pp. 261–266.
38. Lin, M., Popov, A.A., and McWilliam, S., 2004, Stability and performance studies of driver-vehicle systems with electronic chassis control. *Vehicle System Dynamics Supplement*, **41**, 477–486.
39. Lin, M., Popov, A.A., and McWilliam, S., 2003, Sensitivity analysis of driver characteristics in driver-vehicle handling studies, In: L. Dorn (Ed.) *Driver Behaviour and Training* (Aldershot, Hampshire, London, U.K.: Ashgate Publishing Ltd.), (ISBN 0754638359).
40. Johannsen, G., 1993, *Mensch-Maschine-Systeme* (Berlin, Germany: Springer Verlag), (ISBN 3–540–56152–8), (in German).
41. Weir, D.H., DiMarco, R.J., and McRuer, D.T., 1977, *Evaluation and Correlation of Driver/Vehicle Data. Vol. I. Summary Report* (Washington, DC: National Highway Traffic Safety Administration), (DOT-HS-803–246).
42. Hess, R.A. and Modjtahedzadeh, A., 1990, A control theoretic model of driver steering behavior. *IEEE Control Systems Magazine*, **10**(5), 3–8.
43. Garrott, W.R., Wilson, D.L., and Scott, R.A., 1982, Closed loop automobile maneuvers using describing function models. *Society of Automotive Engineers (SAE)*, 820306.
44. Festa, P., Mangialardi, L., and Mantriota, G., 2003, Dynamic behaviour of a closed loop driver-car-caravan system. *International Journal of Vehicle Design*, **31**(4), 408–426.
45. Mitschke, M. and Ahring, E., 1994, Control loop for driver-vehicle with four wheel steering. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 1994*, Tsukuba-shi, Japan.
46. Plöchl, M. and Lugner, P., 1999, A 3-level driver model and its application to driving simulation. *Vehicle System Dynamics Supplement*, **33**, 71–82.
47. Edelmann, J., Plöchl, M., Reinalter, W., and Tieber, W., 2007, A passenger car driver model for higher lateral accelerations. *Vehicle System Dynamics*, **45**(12), 1117–1129.
48. Savkoor, A.R. and Ausejo, S., 1999, Analysis of driver's steering and speed control strategies in curve negotiation. *Vehicle System Dynamics Supplement*, **33**, 94–109.
49. Donges, E., 1978, A two-level model of driver steering behaviour. *Human Factors*, **20**(6), 691–707.
50. Mitschke, M., 1989, Anticipatory steering in a driver-vehicle control loop. *Vehicle System Dynamics Supplement*, **18**, 405–413.
51. Edelmann, J., Plöchl, M., and Lugner, P., 2007, A driver model for vehicle system dynamics simulation. *Proceedings of the European Automotive Congress "Automobile for the Future" 2007*, Budapest, Hungary.
52. Bösch, P., 1991, Der Fahrer als Regler. PhD thesis, Vienna University of Technology, Vienna, Austria, (in German).
53. MacAdam, C.C., 1981, Application of an optimal preview control for simulation of closed-loop automobile driving. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-11**(6), 393–399.
54. Ungoren, A.Y. and Peng, H., 2005, An adaptive lateral preview driver model. *Vehicle System Dynamics*, **43**(4), 245–259.
55. Peng, H., 2002, Evaluation of driver assistance systems—A human centered approach. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2002*, Hiroshima, Japan.
56. Guo, K., Ding, H., Zhang, J., Lu, J., and Wang, R., 2004, Development of a longitudinal and lateral driver model for autonomous vehicle control. *International Journal of Vehicle Design*, **36**(1), 50–65.
57. Guo, K., Cheng, Y., and Ding, H., 2004, Analytical method for modeling driver in vehicle directional control. *Vehicle System Dynamics Supplement*, **41**, 401–410.

58. Guo, K., Ding, H., Wan, F., Lv, J., and Cao, J., 2006, A unified analytical method for lateral and longitudinal combined driver model. *Proceedings of the International Symposium on Advanced Vehicle Control* (*AVEC*) *2006*, Taipei, Taiwan.
59. Sharp, R.S., Casanova, D., and Symonds, P., 2000, A mathematical model for driver steering control, with design, tuning and performance results. *Vehicle System Dynamics*, **33**, 289–326.
60. Kroll, C.V., 1971, Preview-predictor model of driver behavior in emergency situations. *Driving Simulation* (Washington, DC: Highway Research Board), vol. 364, pp. 16–26, (ISBN 0309019796).
61. Sharp, R.S. and Valtetsiotis, V., 2001, Optimal preview car steering control. *Vehicle System Dynamics Supplement*, **35**, 101–117.
62. Sharp, R.S., 2005, Driver steering control and a new perspective on car handling qualities. *Proceedings of IMechE, Part C: Journal of Mechanical Engineering Science*, **219**, 1041–1051.
63. Cole, D.J., Pick, A.J., and Odhams, A.M.C., 2006, Predictive and linear quadratic methods for potential application to modelling driver steering control. *Vehicle System Dynamics*, **44**(3), 259–284.
64. Frezza, R., Saccon, A., and Bachet, D., 2003, SmartDriver: A sensor based model of a car driver for virtual product development. *Proceedings of the 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, **1**, 366–370.
65. Frezza, R. and Minen, D., 2005, A new model based driver program for virtual cars in closed loop simulation. *Presented at IAVSD-Symposium 2005*, Milano, Italy.
66. Mayr, R. and Freund, E., 1992, On the Design of nonlinear path control in automated vehicle guidance. *Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems*, **1**, 613–620.
67. Vögel, M., 1997, Fahrbahnmodellierung und Kursregelung für ein echtzeitfähiges Fahrdynamikprogramm. Diploma-thesis, Munich University of Technology, Munich, Germany, (in German).
68. Vögel, M., von Stryk, O., Bulirsch, R., Wolter, T.-M., and Chucholowski, C., 2003, An optimal control approach to real-time vehicle guidance, In: W. Jäger and H.-J. Krebs (Eds.), *Mathematics—Key Technology for the Future* (pp. 84–102) (Berlin, Germany: Springer Verlag).
69. Irmscher, M. and Ehmann, M., 2004, Driver classification using ve-Dyna advanced driver. *Society of Automotive Engineers (SAE)*, **1877**, 129–134, 2004–01–0451.
70. Butz, T. and von Stryk, O., 2005, Optimal control based modeling of vehicle driver properties. *Society of Automotive Engineers (SAE)*, pp. 7, 2005-01-0420.
71. Zadeh, G.A., Fahim, A., and El-Gindy, M., 1997, Neural network and fuzzy logic applications to vehicle systems: Literature survey. *International Journal of Vehicle Design*, **18**(2), 132–193.
72. Hessburg, T. and Tomizuka, M., 1994, Fuzzy logic control for lateral vehicle guidance. *Control Systems Magazine*, **14**, 55–63.
73. Zeyada, Y., El-Beheiry, E., El-Arabi, M., and Karnopp, D., 2000, Driver modeling using fuzzy logic controls for human-in-the-loop vehicle simulations, In: Hassan, M.F., Megahed, S.M. (Eds.), *Current Advances in Mechanical Design and Production VII* (Oxford, U.K.: Elsevier Science Ltd.).
74. Xi, G. and Qun, Y., 1994, Driver-vehicle-environment closed-loop simulation of handling and stability using fuzzy control theory. *Vehicle System Dynamics Supplement*, **23**, 172–183.
75. Kageyama, I. and Pacejka, H.B., 1991, On a new driver model with fuzzy control. *Vehicle System Dynamics Supplement*, **20**, 314–324.
76. Kramer, U. and Rohr, G., 1982, A fuzzy model of driver behaviour: Computer simulation and experimental results. *Proceedings of IFAC Symposium on Analysis, Design, and Evaluation of Man-Machine Systems*, Baden-Baden, Wurttemberg, Germany, pp. 31–35.
77. Di Puccio, F., Forte, P., Guiggiani, M., and Rotti, D., 2003, Modelling and simulation of driver's control on 4-wheeled vehicle dynamics. *CD-Proceedings of 4th International Conference on Control and Diagnostics in Automotive Applications*, Sestri Levante, Italy.
78. James, D.J.G., Böhringer, F., Burnham, K.J., and Copp, D.G., 2002, Adaptive driver model using neural network. *Proceedings of the Artificial Life and Robotics*, (ISSN 1433–5298), **7**, 170–176.
79. Chen, L.K. and Ulsoy, A.G., 2001, Identification of a driver steering model, and model uncertainty, from driving simulator data. *ASME Journal of Dynamic Systems, Measurement, and Control*, **123**, 623–629.
80. Chen, L.K. and Ulsoy, A.G., 2000, Identification of a nonlinear driver model via NARMAX modeling. *Proceedings of the American Control Conference*, Chicago, IL, pp. 2533–2537.
81. Chen, L.K. and Hu, D.J., 2006, Adaptive control with respect to driver model uncertainty. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2006*, Taipei, Taiwan.
82. Chen, L.K. and Lin, H.T., 2006, On-line driver steering model estimations. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2006*, Taipei, Taiwan.

83. Neculau, M., Jürgensohn, T., and Willumeit, H.-P., 1990, Model of the driver lateral control behaviour and scanning pattern. *Proceedings of the 9th European Annual Conference on Human Decision Making and Manual Control*, Ispra, Italy, pp. 275–282.
84. Reddy, R.N. and Ellis, J.R., 1981, Contribution to the simulation of driver-vehicle-road system. *Society of Automotive Engineers (SAE)*, pp. 24, 810513.
85. Wenzel, T.A., Burnham, K.J., Williams, R.A., and Blundell, M.V., 2005, Closed-loop driver/vehicle model for automotive control. *Proceedings of the 18th International Conference on Systems Engineering*, Washington, DC.
86. Riepl, A., Reinalter, W., and Fruhmann, G., 2004, Rough road simulation with tire model RMOD-K and FTire. *Vehicle System Dynamics Supplement*, **41**, 734–743.
87. Lugner, P. and Plöchl, M., 1995, Additional 4WS and driver interaction. *Vehicle System Dynamics*, **24**, 639–658.
88. Plöchl, M. and Lugner, P., 1999, Passenger car and passenger car-trailer—Different tasks for the driver. *JSAE Review*, **20**, 543–548.
89. Stribersky, A., Fancher, P.S., MacAdam, Ch.C., and Sayers, M.W., 1989, On nonlinear oscillations in road trains at high forward speeds. *Vehicle System Dynamics Supplement*, **18**, 552–565.
90. Harada, H. and Iwasaki, T., 1994, Stability criteria and objective evaluation of a driver-vehicle system for driving in lane change and against crosswind. *Vehicle System Dynamics Supplement*, **23**, 197–208.
91. Kehtarnavaz, N. and Sohn, W., 1991, Steering control of autonomous vehicles by neural networks. *Proceedings of the American Control Conference*, IEEE cat 91CH2939-7, Boston, MA, pp. 3096–3101.
92. Pomerleau, D., 1991, Efficient training of artificial neural networks for autonomous navigation. *Neural Computation*, **3**(1), 88–97.
93. Pomerleau, D., 1993, Knowledge-based training of artificial neural networks for autonomous robot driving, In: Connell, J., Mahadevan, S. (Eds.), *Robot Learning* (Boston, MA: Kluwer Academic Publishing).
94. Lubin, J.M., Huber, E.C., Gilbert, S.A., and Kornhauser, A.L., 1992, Analysis of a neural network lateral controller for an autonomous road vehicle. *Society of Automotive Engineers (SAE)*, pp. 21, 921561.
95. Neusser, S., Hoefflinger, B., and Nijhuis, J., 1991, A case study in car control by neural networks. *Proceedings of the ISATA International Symposium*, Florence, Italy, 1991, pp. 607–614.
96. Neusser, S., Nijhuis, J., Spaanenburg, L., Höfflinger, B., Franke, U., and Fritz, H., 1993, Neurocontrol for lateral vehicle guidance. *IEEE Micro*, **13**(1), 57–66.
97. Rivals, I., Canas, D., Personnaz, L., and Dreyfus, G., 1994, Modeling and control of mobile robots and intelligent vehicles by neural networks. *Proceedings of the Intelligent Vehicles Symposium*, Paris, France, pp. 137–142.
98. Fritz, H., 1996, Model-based neural distance control for autonomous road vehicles. *Proceedings of the Intelligent Vehicles Symposium*, Tokyo, Japan, pp. 29–34.
99. Kehtarnavaz, N., Griswold, N., Miller, K., and Lescoe, P., 1998, A transportable neural-network approach to autonomous vehicle following. *IEEE Transactions on Vehicular Technology*, **47**(2), 694–702.
100. Kamada, H. and Yoshida, M., 1992, A visual control system using image processing and fuzzy theory. *Vision-Based Vehicle Guidance* (New York: Springer Verlag), (ISBN O-387–97553–5).
101. Hessburg, T. and Tomizuka, M., 1995, Fuzzy logic control for lane change maneuvers in lateral vehicle guidance, In: *California Partners for Advanced Transit and Highways (PATH). Working Papers:* Paper UCB-ITS-PWP-95–13.
102. Pasquier, M., Quek, C., Tung, W.L., Chen, D., and Yep, T.M., 2002, Fuzzylot II: A novel soft computing approach to the realisation of autonomous driving manoeuvres for intelligent vehicles. *Proceedings of International Conference on Control, Automation, Robotics and Vision*, **2**, 746–751.
103. Kraiss, K.F. and Küttelwesch, H., 1990, Teaching neural networks to guide a vehicle through an obstacle course by emulating a human teacher. *Proceedings of International Joint Conference on Neural Networks*, San Diego, CA, 333–337.
104. Kraiss, K.F. and Küttelwesch, H., 1992, Identification and application of neural operator models in a car driving situation. *Proceedings of the 5th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, The Hague, The Netherlands 121–126.
105. Fujioka, T., Nomura, K., Takubo, N., and Chung, D.H., 1991, Driver model using neural network system. *Proceedings of 6th International Pacific Conference on Automotive Engineering*, Seoul, South Korea, Vol. 2, pp. 931–938.
106. MacAdam, Ch.C. and Johnson, G.E., 1996, Application of elementary neural networks and preview sensors for representing driver steering control behavior. *Vehicle System Dynamics*, **25**, 3–30.
107. Lin, Y., Tang, P., Zhang, W.J., and Yu, Q., 2005, Artificial neural network modelling of driver handling behaviour in a driver-vehicle-environment system. *International Journal of Vehicle Design*, **37**(1), 24–45.

108. MacAdam, Ch.C., Bareket, Z., Fancher, P., and Ervin, R., 1998, Using neural networks to identify driving style and headway control behavior of drivers. *Vehicle System Dynamics Supplement*, **28**, 143–160.
109. Kageyama, I., Arai, A., and Nomura, T., 1999, An analysis of driver's control algorithm using neural network modeling. *Vehicle System Dynamics Supplement*, **33**, 122–130.
110. Kuriyagawa, Y., Im, H.-E., Kageyama, I., and Onishi, S., 2002, A research on analytical method of driver-vehicle-environment system for construction of intelligent driver support system. *Vehicle System Dynamics*, **37**(5), 339–358.
111. Koike, Y. and Doya, K., 1999, Multiple state estimation reinforcement learning for driving model—Driver model of automobile. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, **5**, 504–509.
112. Koike, Y. and Doya, K., 2003, Driver model based on reinforced learning with multiple-step state estimation. *Electronics and Communications in Japan, Part 3*, **86**(10), 85–95.
113. Wolf, J.D. and Barrett, M.F. (1978), Driver-vehicle effectiveness model, Vol. I: final report, Vol. II: appendices. (Report No. DOT/HS 804337/804338) (Washington, DC: National Highway Traffic Safety Administration).
114. Hayhoe, G.F., 1979, A driver model based on the cerebellar model articulation controller. *Vehicle System Dynamics*, **8**, 49–72.
115. Carson, J.M. and Wierwille, W.W., 1978, Development of a strategy model of the driver in lane keeping. *Vehicle System Dynamics*, **7**, 233–253.
116. Crossman, E.R.F.W. and Szostak, H., 1968, Man-machine models for car steering. *Proceedings of the 4th Annual Conference on Manual Control*, NASA Technical Report SP-192, Berkeley, CA, pp. 171–195.
117. Jürgensohn, T. and Willumeit, H.-P., 1997, Modellierung von Mikro-Steuerstrategien mit Hilfe künstlicher neuronaler Netze. *Automobiltechnische Zeitschrift (ATZ)*, **99**(6), 348–352, (in German).
118. Cheng, B. and Fujioka, T., 1997, A hierarchical driver model. *Proceedings of IEEE Conference on Intelligent Transportation Systems*, Boston, MA, pp. 960–965.
119. Song, B., Delorme, D., and VanderWerf, J., 2000, Cognitive and hybrid model of human driver. *Proceedings of the Intelligent Vehicles Symposium*, Dearborn, MI, pp. 1–6.
120. Salvucci, D.D., 2006, Modeling driver behavior in a cognitive architecture. *Human Factors*, **48**, 362–380.
121. Salvucci, D.D. and Gray, R., 2004, A two-point visual control model of steering. *Perception*, **33**(10), 1233–1248.
122. Kiencke, U., Majjad, R., and Kramer, S., 1999, Modeling and performance analysis of a hybrid driver model. *Control Engineering Practice*, **7**, 985–991.
123. Wickens, C.D., 1984, *Engineering Psychology and Human Performance* (London, U.K.: C.E. Merrill Publishing Company).
124. Liu, Y., 1994, A queueing network model of human performance of concurrent spatial and verbal tasks. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, **3**, 2761–2766.
125. Kiencke, U. and Nielsen, L., 2005, *Automotive Control Systems* (Berlin, Germany: Springer Verlag), (ISBN 3–540–23139–0).
126. Lauffenburger, J.-Ph., Basset, M., and Gissinger, G.L., 2005, Design of a coupled longitudinal-lateral trajectographic driver model. *Proceedings of the 16th IFAC World Congress 2005*, Prague, Czech Republic.
127. Kim, J.-H., Kim, Y.W., and Hwang, D.H., 2005, Modeling of human driving behavior based on piecewise linear model. *Automatika*, **46**(1–2), 29–37.
128. Keen, S.D. and Cole, D.J., 2006, Steering control using model predictive control and multiple internal models. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2006*, Taipei, Taiwan.
129. Nagai, M., Onda, M., and Katagiri, T., 1997, Application of genetic algorithm to analysis of driver's behavior in collision avoidance. *International Journal of Vehicle Design*, **18**(6), 626–638.
130. Li, L. and Wang, F.-Y., 2002, Vehicle trajectory generation for optimal driving guidance. *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, Singapore 231–235.
131. Li, L. and Wang, F.-Y., 2003, An integrated design framework for driver/passenger-oriented trajectory planning. *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, **2**, 1764–1769.
132. Prokop, G., 2001, Modeling human vehicle driving by model predictive online optimization. *Vehicle System Dynamics*, **35**(1), 19–53.
133. Preusse, Ch., 2001, A driver model for online control of virtual cars. *Proceedings of the IEEE International Conference on Control Applications*, 1174–1178.
134. Preuße, Ch., Keller, H., and Hunt, K.J., 2001, Fahrzeugführung durch ein Fahrermodell. *Automatisierungstechnik*, **49**, 540–546, (in German).

135. Lauffenburger, J.-Ph., Basset, M., Coffin, F., and Gissinger, G.L., 2003, Driver-aid system using path-planning for lateral vehicle control. *Control Engineering Practice*, **11**, 217–231.
136. Hoang, H.T., Lauffenburger, J.-Ph., Basset, M., and Thanh, V.D., 2005, An improvement of the driver trajectory model for path-planning. *Proceedings of the International Symposium on Electrical & Electronics Engineering 2005*, HCM City, Vietnam.
137. Odhams, A.M.C. and Cole, D.J., 2004, Models of driver speed choice in curves. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2004*, Arnhems, the Netherlands, pp. 439–444.
138. Sentouh, C., Glaser, S., and Mammar, S., 2006, Advanced vehicle-infrastructure-driver speed profile for road departure accident prevention. *Vehicle System Dynamics Supplement*, **44**, 612–623.
139. Casanova, D., Sharp, R.S., and Symonds, P., 2000, Minimum time manoeuvring: The significance of yaw inertia. *Vehicle System Dynamics*, **34**(2), 77–115.
140. Casanova, D., Sharp, R.S., and Symonds, P., 2001, Sensitivity to mass variations of the fastest possible lap of a formula one car. *Vehicle System Dynamics Supplement*, **35**, 119–134.
141. Magerl, F., 2003, Ein Fahrermodell für die Simulation des ISO-Spurwechsels, PhD-thesis, Vienna University of Technology, Vienna, Austria, (in German).
142. Boyer, F. and Lamiraux, F., 2006, Trajectory deformation applied to kinodynamic motion planning for a realistic car model. *Proceedings of IEEE Conference on Robotics and Automation*, Orlando, FL, pp. 487–492.
143. Bernard, J., Gruening, J., and Hoffmeister, K., 1998, Evaluation of vehicle/driver performance using genetic algorithms. *Society of Automotive Engineers (SAE)*, pp. 13, 980227.
144. Horiuchi, S., Okada, K., and Nohtomi, S., 2004, Optimum steering and braking control strategies in obstacle avoidance maneuvers. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC), 2004*, Arnhems, the Netherlands, pp. 619–624.
145. Fujioka, T., Kouge, Y., Fukuba, H., Adachi, T., and Yoshioka, T., 2004, Analysis of driving behavior on expert driver in double lane change through optimization. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC), 2004*, Arnhems, the Netherlands, pp. 421–426.
146. Soma, H. and Hiramatsu, K., 1994, Application of AR-method to dynamic identification of driver-vehicle system in lane change. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 1994*, Tsukuba-shi, Japan, pp. 22–27.
147. Sonada, T. and Makita, M., 2006, Analysis of driver maneuver with optimizing driver model for double lane change. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2006*, Taipei, Taiwan.
148. El Hajjaji, A. and Ouladsine, M., 2001, Modeling human vehicle driving by fuzzy logic for standardized ISO double lane change maneuver. *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, Bordeaux and Paris, France 499–503.
149. Gordon, T.J. and Best, M.C., 2006, On the synthesis of driver inputs for the simulation of closed-loop handling manoeuvres. *International Journal of Vehicle Design*, **40**(1/2/3), 52–76.
150. Gordon, T.J. and Best, M.C., 2004, Comparison of optimal driving policies for limit handling manoeuvres. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 2004*, Arnhems, the Netherlands, pp. 399–404.
151. Apetaur, M. and Opicka, F., 1991, Assessment of the driver's effort in typical driving manoeuvres for different vehicle configurations and managements. *Vehicle System Dynamics Supplement*, **20**, 42–56.
152. Horiuchi, S. and Yuhara, N., 2000, An analytical approach to the prediction of handling qualities of vehicles with advanced steering control system using multi-input driver model. *Dynamic Systems, Measurement, and Control*, **122**(3), 490–497.
153. Limpibunterng, T. and Fujioka, T., 2004, Bilateral driver model for steer-by-wire controller design. *Vehicle System Dynamics Supplement*, **41**, 381–390.
154. Yang, X., Rakheja, S., and Stiharu, I., 2001, Adapting an articulated vehicle to its drivers. *Journal of Mechanical Design*, **123**(1), 132–140.
155. Yang, X., Rakheja, S., and Stiharu, I., 2002, Structure of driver model for articulated vehicles. *International Journal of Vehicle Design, Heavy Vehicle Systems*, **9**(1), 27–51.
156. Chai, Y.W., Abe, Y., Kano, Y., and Abe, M., 2006, A study on adaptation of SBW parameters to individual driver's steer characteristics for improved driver-vehicle system performance. *Vehicle System Dynamics Supplement*, **44**, 874–882.
157. Fujiwara, Y., Fujihira, T., Ishiwa, S., and Adachi, S., 2004, Control design of driver support system using multiple driver models. *Proceedings of the SICE 2004 Annual Conference*, **3**, 2443–2448.

158. Furukawa, Y., Takeda, H., Yuhara, N., Chikamori, S., Abe, M., and Sawada, T., 2004, Development of predicting evaluation platform of advanced driver safety assist systems using integrated virtual traffic environments. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, **4**, 4041–4046.
159. Ohno, H., 2000, Analysis and modeling of human driving behaviors using adaptive cruise control. *Proceedings of the 26th Annual Conference of the IEEE Industrial Electronics Society*, **4**, 2803–2808.
160. Feraric, J.P., Kopf, M., and Onken, R., 1992, Statistical versus neural-net approach for driver behaviour description and adaptive warning. *Proceedings of 11th European Annual Conference on Human Decision Making and Manual Control*, 15p.
161. Onken, R., 1994, DAISY, an adaptive knowledge-based driver monitoring and warning system. *Proceedings of the Conference on Vehicle Navigation and Information Systems*, Berkeley, CA, pp. 3–10.
162. Onken, R. and Feraric, J.P., 1997, Adaptation to the driver as part of a driver monitoring and warning system. *Accident Analysis and Prevention*, **29**(4), 507–513.
163. Liu, A. and Pentland, A., 1997, Towards real-time recognition of driver intentions. *Proceedings of the IEEE Conference on Intelligent Transportation System*, Cambridge, MA, pp. 236–241.
164. Miyazaki, T., Kodama, T., Furuhashi, T., and Ohno, H., 2001, Modeling of human behaviors in real driving situations. *Proceedings of the IEEE Conference on Intelligent Transportation System*, Oakland, CA, 643–646.
165. Torkkola, K., Gardner, M., Wood, C., Schreiner, C., Massey, N., Leivian, B., Summers, J., and Venkatesan, S., 2005, Toward modeling and classification of naturalistic driving. *Proceedings of the Intelligent Vehicles Symposium*, Tempe, AZ, pp. 638–643.
166. Bouslimi, W., Kassaagi, M., Lourdeaux, D., and Fuchs, P., 2005, Augmented naive Bayesian network for driver behavior modeling. *Proceedings of the Intelligent Vehicles Symposium*, Berlin, Germany, pp. 236–242.
167. Dagli, I., Brost, M., and Breuel, G., 2002, Action recognition and prediction for driver assistance systems using dynamic belief networks. *Proceedings of Conference on Agent Technologies, Infrastructures, Tools, and Applications for E-Services*, 179–194.
168. Ohashi, L., Yamaguchi, T., and Tamai, I., 2004, Humane automotive system using driver intention recognition. *Proceedings of the SICE 2004 Annual Conference*, **2**, 1164–1167.
169. Kim, J.-H., Hayakawa, S., Suzuki, T., Hayashi, K., Okuma, S., Tsuchida, N., Shimizu, M., and Kido, S., 2005, Modeling of driver's collision avoidance maneuver based on controller switching model. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-35**(6), 1131–1143.
170. Chan, Ch.-Y. and Tan, H.-S., 2001, Feasibility analysis of steering control as a driver-assistance function in collision situations. *IEEE Transactions on Intelligent Transportation Systems*, **2**(1), 1–9.
171. Fournier, L., 1994, Learning capabilities for improving automatic transmission control. *Proceedings of the Intelligent Vehicles Symposium*, Paris, France, pp. 455–460.
172. Brackstone, M. and McDonald, M., 1999, Car-following: A historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, **2**(4), 181–196.
173. Reuschel, A., 1950, Fahrzeugbewegungen in der Kolonne. *(Austrian) Ingenieur-Archiv*, **IV**(3–4), pp. 193–215, (in German).
174. Dreyer, W. and Mitschke, M., 1984, Regelkreis Fahrer/Kraftfahrzeug und Kolonnenfahrt. *Automobil Industrie*, **1/84**, pp. 95–101, (in German).
175. Mulder, M., Muldner, M., van Paassen, M.M.R., and Abbink, D.A., 2005, Identification of driver car-following behaviour. *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, **3**, 2905–2910.
176. Pipes, L.A., 1953, An operational analysis of traffic dynamics. *Journal of Applied Physics*, **24**(3), 274–281.
177. Chandler, R.E., Herman, R., and Montroll, E.W., 1958, Traffic dynamics: Studies in car following. *Operations Research*, **6**(2), 165–184.
178. Bose, A. and Ioannou, P., 1999, Analysis of traffic flow with mixed manual and semi-automated vehicles. *Proceedings of the American Control Conference*, San Diego, CA, pp. 2173–2177.
179. Denos, C., Gazis, R.H., and Rothery, R.W., 1961, Nonlinear follow-the-leader models of traffic flow. *Operations Research*, **9**(4), 545–567.
180. Chung, S.B., Song, K.H., Hong, S.Y., and Kho, S.Y., 2005, Development of sensitivity term in car-following model considering practical driving behavior of preventing rear end collision. *Journal of Eastern Asia Society for Transportation Studies*, **6**, 1354–1367.
181. Helly, W., 1959, Simulation of bottlenecks in single-lane traffic flow. *Proceedings of the Symposium on Theory of Traffic Flow*, Research Laboratories, General Motors, Amsterdam, the Netherlands, pp. 207–238.

182. Ma, X., 2006, A neural fuzzy framework for modeling car-following behavior. *Proceedings of the IEEE International Conference on System, Man and Cybernetics*, Tainan, Taiwan, 6p.
183. Tyler, J.S., Jr., 1964, The characteristics of model-following systems as synthesized by optimal control. *IEEE Transactions on Automatic Control*, **AC-9**(4), 485–498.
184. Burnham, G.O., Seo, J., and Bekey, G.A., 1944, Identification of human driver models in car following. *IEEE Transactions on Automatic Control*, **AC-19**(6), 911–915.
185. Gipps, P.G, 1981, A behavioural car-following model for computer simulation. *Transportation Research B*, **15B**, 105–111.
186. Kometani, E. and Sasaki, T., 1959, Dynamic behaviour of traffic with a non-linear spacing-speed relationship. *Proceedings of the Symposium on Theory of Traffic Flow*, Amsterdam, the Netherlands, pp. 105–119.
187. Bando, M., Hasebe, K., Nakanishi, K., Nakayama, A., Shibata, A., and Sugiyama, Y., 1995, Phenomenological study of dynamical model of traffic flow. *Journal de Physique I France*, **5**, 1389–1399.
188. Newell, G.F., 1961, Nonlinear effects in the dynamics of car following. *Operations Research*, **9**(2), 209–229.
189. Lee, K. and Peng, H., 2004, Identification and verification of a longitudinal human driving model for collision warning and avoidance systems. *International Journal of Vehicle Autonomous Systems*, **2**(1–2), 3–17.
190. Bareket, Z., Fancher, P.S., Peng, H., and Lee, K., 2003, Methodology for assessing adaptive cruise control behavior. *IEEE Transactions on Intelligent Transportation Systems*, **4**(3), 123–131.
191. Weng, Y. and Wu, T., 2002, Car-following models of vehicular traffic. *Journal of Zhejiang University Science*, **3**(4), 412–417.
192. Boer, E.R., 1999, Car following from the driver's perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, **2**(4), 201–206.
193. Goodrich, M.A., Boer, E.R., and Inoue, H., 1998, Brake initiation and braking dynamics: A human-centered study of desired ACC characteristics. Technical Report TR 98–5. Cambridge Basic Research, Nissan Research and Development.
194. Brackstone, M. and McDonald, M., 1999, What is the answer? And come to that, what are the questions? *Transportation Research Part F: Traffic Psychology and Behaviour*, **2**(4), 221–224.
195. Suzuki, H., Ranjitkar, P., Nakatsuji, T., and Takeichi, Y., 2005, An extended car-following model combined with a driver model. *Journal of Eastern Asia Society for Transportation Studies*, **6**, 1545–1556.
196. Ranjitkar, P., Nakatsuji, T., and Kawamura, A., 2005, Experimental analysis of car-following dynamics and traffic stability, *Traffic Flow Theory 2005* (pp. 22–32) (Washington, DC: Transportation Research Board), (ISBN 0309094089).
197. Panwai, S. and Hussein, D., 2005, Comparative evaluation of microscopic car-following behavior. *IEEE Transactions on Intelligent Transportation Systems*, **6**(3), 314–325.
198. Bonsall, P.W., Liu, R., and Young, W., 2005, Modelling safety-related driving behavior—The impact of parameter values. *Transportation Research Part A: Policy and Practice*, **39**, 425–444.
199. Kerner, B.S. and Klenov, S.L., 2006, Deterministic microscopic three-phase traffic flow models. *Journal of Physics A: Mathematical and General*, **39**(8), 1775–1809.
200. Iwaki, R., Kaneko, T., and Kageyama, I., 2006, Study on the longitudinal control analysis of a driver of a heavy-duty vehicle following another vehicle. *Vehicle System Dynamics Supplement*, **44**, 216–229.
201. Peeta, S., Zhou, W., and Zhang, P., 2004, Modeling and mitigation of car-truck interactions on freeways. *Transportation Research Record No. 1899*, 117–126.
202. Halati, A., Torres, J.F., and Cohen, S.L., 1991, FRESIM-Freeway simulation model. *70th Transportation Research Board Annual Meeting*, Paper No. 9102202.
203. Jiang, R. and Wu, Q.-S., 2007, The night driving behavior in a car-following model. *Physica A*, **375**(1), 297–306.
204. Suda, Y., Takahashi, Y., Kuwahara, M., Tanaka, S., Ikeuchi, K., Kagesawa, M., Shraishi, T., Onuki, M., Honda, K., and Kano, M., 2005, Development of universal driving simulator with interactive traffic environment. *Proceedings of the Intelligent Vehicles Symposium*, Las Vegas, NV, 743–746.
205. An, P.E. and Harris, C.J., 1996, An intelligent driver warning system for vehicle collision avoidance. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, **26**(2), 254–261.
206. Wang, L., Liu, X., and Zhong, X., 2006, A car-following model based on artificial neural networks in urban expressway sections. *85th Transportation Research Board Annual Meeting*, Compendium of Papers CD-ROM, 26p.
207. Das, S., Bowles, B.A., Houghland, Ch.R., Hunn, S.J., and Zhang, Y.L., 1999, A knowledge based model of traffic behavior in freeways. *Proceedings of the 1999 ACM Symposium on Applied Computing*, New York, pp. 14–18.

208. Lukanin, V.N., Buslaev, A.P., Novikov, A.V., and Yashina, M.V., 2003, Traffic flows modelling and the evaluation of energy-ecological parameters. Part I and II. *International Journal of Vehicle Design*, **33**(4), 381–399 and 400–421.
209. Leutzbach, W., 1988, *Introduction to the Theory of Traffic Flow* (Berlin, Germany: Springer Verlag).
210. Leutzbach, W. and Wiedemann, R., (1986) Development and applications of traffic simulation models at the Karlsruhe Institut für Verkehrswesen. *Traffic Engineering and Control*, **27**(5), 270–278.
211. Ludmann, J. and Neunzig, D., 1996, The effectivity of new traffic technologies in suburban areas—An analysis with the microscopic traffic simulator PELOPS. *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC) 1996*, Aachen, Germany, pp. 1401–1432.
212. Kumamoto, H., Nishi, K., Tenmoku, K., and Shimoura, H., 1995, Rule based cognitive animation simulator for current lane and lane change drivers. *Proceedings of the 2nd World Congress on ATT*, Yokohama, Japan, pp. 1746–1752.
213. Barton, J.E., Cohn, T.E., and Tomizuka, M., 2006, Towards a complete human driver model: The effect of vision on driving performance. *Proceedings of the American Control Conference*, Minneapolis, MN, pp. 2591–2598.
214. Michaels, R.M., 1963, Perceptual factors in car following. *Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow*, Paris, France, pp. 44–59.
215. Sauer, C.W., Andersen, G.J., and Saidpour, A., 2003, Car following by optical parameters. *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Park City, UT, pp. 158–162.

Part VII

Intelligent Vehicle Systems

39 Automatic Lateral Vehicle Control

Huei Peng

CONTENTS

39.1 Introduction 1525
39.2 Lane Keeping 1526
39.2.1 Background and Literature Review 1526
39.2.2 Sensing Systems 1527
39.2.3 Advanced Control Concepts 1528
39.2.4 Lane Keeping Assist 1530
39.2.5 Vehicle State Estimation 1531
39.2.5.1 Kalman Filter Method 1532
39.2.5.2 Transfer Function Method 1533
39.2.5.3 State-Space Method 1534
39.2.5.4 Kinematics Method 1536
39.2.5.5 Road Bank Angle Estimation 1537
39.3 Precise Lateral Maneuver 1539
39.3.1 Automated Buses 1540
39.3.2 Parking Assist and Automated Valet 1540
39.4 Lane Departure Warning and Prevention 1541
39.4.1 Calculation of Time-to-Lane Crossing 1542
39.4.2 Estimation of Tire Cornering Stiffness 1543
39.4.3 Warning and Control Algorithms 1544
39.5 Conclusions 1545
Acknowledgments 1545
References 1545

39.1 INTRODUCTION

The main purpose of this chapter is to provide an overview of the dynamics and control issues related to the handling of vehicles driving on paved roadways. The division of this section of the handbook ("Intelligent Vehicle Systems") into a longitudinal and a lateral chapter is somewhat artificial—there are very close interactions between vehicle lateral and longitudinal dynamics. Vehicle controls coordinating lateral/yaw and longitudinal motions simultaneously are obviously beneficial and are even necessary in tasks such as merge assistance for automated highway systems. In addition, there will be unavoidable overlap with several chapters of Part V of this handbook. Here is the basic idea in delineating the material to be covered in this chapter: lane position. Control of horizontal vehicle motion could include important functionalities such as lateral stability to prevent spinout or plow-out, which is covered very nicely in Part V. For this chapter, it is assumed that the lane position is detected and used in vehicle motion regulation. If this is not the case, the related topics will not be included. We will try to provide a self-contained introduction

of the current state-of-the-art development in the area of intelligent vehicles related to vehicle handling and lane keeping functionalities. We will also highlight the issues related to the coordination between the vehicle and the roadway infrastructure when appropriate.

The term "intelligent vehicle and highway systems" was created in the late 1980s in the United States to encompass an array of applications related to the deployment of advanced information, electronics, and control technologies to improve the efficiency and safety of highway vehicles. This umbrella term covers a wide range of technologies, and the subarea closely related to the material to be covered in this chapter is known as "advanced vehicle control systems" (AVCS). In the early 2000s, a more focused set of efforts was selected and supported within the USDOT's Intelligent Vehicle Initiative (IVI). Identified research challenges related to this chapter include (1) lane keeping function, (2) precision maneuver (e.g., docking of buses or parking in tight spaces), and (3) lane departure warning and prevention. These three areas encompass the majority of the recent progress in the "lateral" direction of intelligent vehicle research. The status of these three function areas will be discussed in the remainder of this chapter.

39.2 LANE KEEPING

39.2.1 Background and Literature Review

Lane keeping systems require the integration of sensing and control systems so that the controlled vehicle automatically follows the detected road markers, which usually delineates a lane of a roadway. We chose not to discuss much about the vehicle lane change functions because of two reasons: (1) Lane change is mainly a decision problem considering traffic flow and relative longitudinal motion of the vehicles in adjacent lanes (in other words, even though the final control maneuver executed is a steering action, the key challenge lies within the longitudinal domain rather than in the lateral domain) and (2) once the decision to make a lane change is made, the lateral control maneuver can be viewed as a special case of lane keeping, with a planned lane position. In other words, once the fundamental sensing and dynamic control problems for lane keeping is solved, the lane change problem is a simple extension of the same problem. Even though there were quite a few papers published recently in the area of lane changing [1–3], we chose not to cover those topics in this chapter.

The automatic lane keeping problem of a vehicle was studied as early as the late 1950s. Gardels [4] addressed the automatic steering control problem using electronic sensors at the General Motors Research. Cardew [5] performed automatic steering tests and achieved position accuracy of ±2 cm on straight roads. These early efforts demonstrated the viability of automatic lane keepings and inspired continuing research in this area. Substantial efforts were made by Fenton and his colleagues through the 1970s [6,7]. They first designed control algorithms using classical control methods [6] and validated their designs by experiments. The classical design approach is adopted due to the fact that digital control technologies were not readily available. Nevertheless, they were able to demonstrate automatic driving on curves with mild lateral accelerations at 0.2–0.3 g. In the late 1980s, the linear–quadratic (LQ) approach was investigated by Fenton et al. [7], and the results were evaluated in a laboratory study where an analog computer was used to simulate the vehicle dynamics. In both [6] and [7], a feedback controller was used to carry out the lane following task, which was treated as a regulation type of control problem. They relied on the disturbance rejection capability of the feedback controller to track curved roads.

The automatic lane keeping function receives a lot of attention under the umbrella of intelligent vehicle and highway system research, starting from the late 1980s and culminated in the late 1990s. Important efforts include the PATH program in the United States [8–11], ASV/AHS in Japan [12,13], and PROMETHEUS/Drive/CARTS programs in Europe [14,15]. These efforts were driven by a surge in the interest and research funding in the area of intelligent vehicles and transportation systems. The PATH program was mainly supported by the California Department of Transportation

(Caltrans) and to a lesser degree by federal and industrial funding. This research program was started in 1987 and now operates at a smaller but still significant level. Japan's ASV project began in 1991 under the leadership of the (former) Ministry of Transport (MOT). The first two phases of the ASV project span from 1991 to 1995 and 1996 to 2000. Under the coordination and guidance of the MOT, participating Japanese automotive companies developed a wide variety of advanced technologies to enhance safety. The Japan AHS project was led by the Ministry of Construction and was carried out jointly with car companies through the public–private partnership "AHSRA." They conducted an AHS demonstration at the Public Works Research Institute (PWRI) in 1995 and on a stretch of public highway in 1996. It is important to point out that the full name for the Japan "AHS" effort is now "Advanced Cruise-Assist Highway System," which seems to be a gentle way to acknowledge the fact that the focus was no longer on fully automated highway systems. This is an obvious departure from the meaning of AHS used in the United States. PROMETHEUS is a research program approved as a EUREKA project and was launched by the European automotive industry in 1986. This project group is then continued through a number of follow-up projects. Automatic steering concepts seem to receive less attention in the European compared with the U.S./Japanese projects. Systems that deal with automatic steering functions, such as those described in [14,15], are of relatively small scale.

A notable achievement of this era, in comparison with the achievement up to the early 1980s, includes sensing systems that are easier to retrofit to existing roads, more reliable and smoother steering controls through advanced control methodologies, and extensions to diverse vehicle types and maneuvers. Key advancements in these areas are explained in the following subsections.

39.2.2 Sensing Systems

To achieve accurate and smooth lane keeping, two types of information are critical: instantaneous vehicle lateral position and orientation and future roadway geometry. Among these variables, the lateral position must be measured, because of vehicle observability requirement. In other words, the latter two variables can be estimated based on the lateral displacement (at the price of time delay), but the opposite is not true. Most of the pre-1980 research prototype systems rely on sensing systems that do not look ahead. Examples include magnetic wires [6,7] and side-looking radars. In other words, these sensing systems focus on obtaining the vehicle lateral position, which can be used to infer vehicle orientation information. However, future road geometry (curvature) cannot be derived. Even when someone takes double derivative of the measured lateral displacement signal, a noisy and impractical approach, what was obtained is the instantaneous curvature error, which is a poor estimation for future road curvature. This fact severely limited the real-world performance of the early downward-looking and side-looking sensing and lane keeping systems, because they essentially drive without looking ahead! Driving while looking at roadways ahead, that is, "previewing," is included in almost all the newer designs in the post-1990 research programs [8–13]. By looking ahead, future road curvature information is used to determine proper steering function while making the overall steering smoother. Previewing is possible by using sensing systems such as vision [16–19], magnetic markers [20,21], and differential GPS plus maps [22,23]. The functioning principles and technical challenges of these sensing systems were amply discussed in the literature and are beyond the scope of this chapter. It is, however, important to summarize the ballpark performance of these sensing systems, their key challenges, and potential for further improvement. The comparison is shown in Table 39.1.

A comprehensive real-world testing and evaluation on lateral positioning and sensing system was done by the Japan AHS-i program [24]. From 1999 to 2001, two promising methods—magnetic markers and passive multiplier radio wave markers—were selected for detailed evaluation. Three new methods were added for evaluations in 2001: pseudo-satellite, direction of arrival (DOA), and dedicated short-range communication (DSRC). The discrete magnetic lane marker system was

TABLE 39.1
Comparison of Vehicle Lateral Position (and Yaw) Sensing System

	Lateral Position Accuracy (cm)	Yaw Angle Accuracy (deg)	Preview Distance and Quality	Reliability	Key Challenges
Vision	2–7 cm	1°–3°	120 m, poor	Poor	Computation, ambient lighting, snow, lane delineation
Magnetic markers	0.5–4.0 cm	0.2°–1.5°	Very long (precoded), good	Excellent	High installation cost
DGPS map	3–10 cm	No direct measurement	Very long (database), good	Good (urban canyons can cause problems)	Accuracy, urban canyon, DGPS stations

found to provide accurate lane position information (4 cm) under varying road surface conditions, weather, road construction, and traffic environment. The positioning accuracy is dependent on the road construction survey and installation accuracy and strength of the magnetic markers. The accuracy (4 cm) reported in [24] is based on markers with intensity of 170 μT with the sensing heads placed 30cm above road surface. This accuracy is somewhat inferior compared to that reported by the California PATH program (0.5 cm, [20]), which is, however, obtained from a short test track. The markers in [20] were installed in specially designed capsules with the markers installed off-center. The markers were then precisely aligned by rotating the markers around the capsule center to achieve precise lateral accuracy. For large-scale installations, much higher installation errors are expected. The high survey/installation cost remains an issue for large-scale implementations of magnetic (and other types of) markers.

The pseudo-satellite technology is designed to augment GPS. Stations at strategically selected locations receive and retransmit GPS signals to cover shadow areas where GPS satellites are known to be blocked, for example, at urban canyons created by high-rise buildings. This technology likely will not provide accuracy beyond that of existing GPS and DGPS techniques. The other two methods (DOA and DSRC) do not seem to be accurate enough for lateral control purposes (lateral position accuracy >1.5 m) and thus are not discussed here.

It is important to point out that the discussion earlier mainly focuses on non vision systems; CCD camera-based vision systems are still a popular choice for automated guided vehicles. Recent examples of vision-based lane tracking systems include [25–28], which represent just a short list of the huge literature in this area. Their inability to deal with snow/ice and the problem of blocked view by adjacent vehicles severely limited their acceptance and practicality for safe day-to-day operations, especially for vehicle control tasks. However, due to the fact that the vision sensor contains rich information useful for other purposes, it will remain a stronger contender, at least as a redundant sensing system.

39.2.3 Advanced Control Concepts

The importance of choosing a sensing system that can look ahead is quite intuitive and can be explained by analyzing the basic control system architecture. Figure 39.1 shows a $2\frac{1}{2}$ degree-of-freedom (DOF) picture of the vehicle road following scenario. The 2 DOF are the vehicle lateral and yaw motions, and the "$\frac{1}{2}$" DOF is the vehicle forward speed, which is assumed to be constant (slow varying), thus treated as a parameter rather than a dynamic state variable in the overall model. It is known that vehicle lateral–yaw–roll dynamics are coupled; thus, this $2\frac{1}{2}$-DOF model assumption

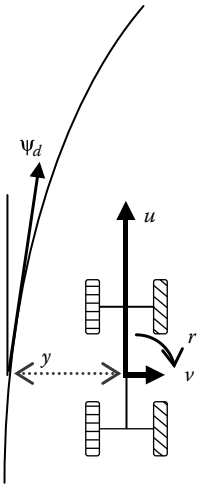


FIGURE 39.1 Road following of a vehicle described by the "bicycle model."

implies that the vehicle roll motion is small and is neglected. In Figure 39.1, u, v, and, r denote vehicle longitudinal speed, lateral speed, and yaw rate, respectively. y is the lateral displacement of the vehicle center of gravity (CG), measured sideways along the y-axis of the vehicle-fixed coordinate system. Integrating vehicle yaw rate, we obtain the vehicle yaw angle (orientation angle) with respect to the earth-fixed coordinate system. If we denote the vehicle yaw angle as $\psi(t)$, it will be possible to define a "desired vehicle yaw angle" $\psi_d(t)$, which is defined as the orientation of the tangential line of the instantaneous road trajectory.

The relationship between $\psi_d(t)$ and the actual road geometric variable, that is, the road radius of curvature, $R(t)$, is $\dot{\psi}_d(t) = r_d(t) = u(t)/R(t)$. The $2\frac{1}{2}$-DOF model can then be described in the state-space form as

$$\frac{d}{dt}\begin{bmatrix} y \\ v \\ \psi - \psi_d \\ r \end{bmatrix} = \begin{bmatrix} 0 & 1 & u & 0 \\ 0 & -\dfrac{C_{\alpha f} + C_{\alpha r}}{mu} & 0 & \dfrac{bC_{\alpha r} - aC_{\alpha f}}{mu} - u \\ 0 & 0 & 0 & 1 \\ 0 & \dfrac{bC_{\alpha r} - aC_{\alpha f}}{I_z u} & 0 & -\dfrac{a^2 C_{\alpha f} + b^2 C_{\alpha r}}{I_z u} \end{bmatrix} \begin{bmatrix} y \\ v \\ \psi - \psi_d \\ r \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{C_{\alpha f}}{m} \\ 0 \\ \dfrac{aC_{\alpha f}}{I_z} \end{bmatrix} \delta + \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} r_d \tag{39.1}$$

where

m is the vehicle mass
I_z is the yaw moment of inertia
$C_{\alpha}f$ and $C_{\alpha r}$ are the front and rear cornering stiffness (per axle)
δ is the front wheel steering angle
a and b are the distance from vehicle CG to front and rear axles, respectively

It can be seen that the road curvature (or radius of curvature) enters this road following dynamic system as a disturbance input. In addition, by examining the input matrices associated with δ and r_d, it is obvious that the steering input does not influence the vehicle dynamics through the same channel as the road curvature input. Therefore, it is impossible to completely cancel the effect of road curvature by using the steering input, that is, perfect control (achieving zero

tracking error) is impossible, even when the available bandwidth of the control system is infinite. What distinguishes this road following dynamics from many other vehicle control applications is the fact that the road curvature is fixed and could be known in advance. Therefore, using "preview" type of control algorithms makes perfect sense and was found to improve the lateral tracking performance significantly [8,20,21]. In fact, using the road curvature in a previewed fashion helps to achieve not only more precise lateral tracking but also smoother ride (reduced peak lateral acceleration). Optimal preview control law based on the linear–quadratic control theory was found to exhibit an out–in–out behavior similar to what a professional driver will do, in order to achieve a trade-off between accurate tracking and smooth ride. The preview control concept lays a solid foundation for the California PATH's lateral control architecture from the late 1980s until the early 2000s [8,22]. Solutions of the preview control algorithm are somewhat mathematically involved and thus were not included in this chapter. Interested readers can refer to [8] for details. It is also worthwhile to point out that the preview control formulation shown in [8] is based on the calculus of variations, and the solution looks more complicated than the optimization results used in the well-known preview driver model developed by MacAdam [29,30], which is incorporated in the CarSim software. The simpler preview solution in the MacAdam formulation works quite well and could be used if a road following is the main concern. The formulation of [8], however, is more flexible and can include more general control objectives in the cost function, for example, ride quality.

39.2.4 Lane Keeping Assist

When the steering control system is designed for "lane keeping assist" (LKA), rather than "automated driving," the control concept, architecture, and algorithm become quite different, even though these two systems share a similar goal (lane keeping). In the LKA case, a human driver is not only in the loop, he or she also has the ultimate authority and responsibility. Therefore, it is necessary to design the system architecture so that the control actions of the automatic system and the human driver can be added in a coherent and robust fashion. In a very fundamental level, one can argue that the best way to harmonically integrate the human and control actions is to use low gains for the control action under normal situations and use higher gains when the driver seems to relinquish the driving task and the vehicle is departing from the lane inadvertently. In [31], it was argued that steering torques from driver/machine can be added together much more easily than steering angle commands. However, since steering torque acts upstream of the (significantly nonlinear) steering system, which commonly includes an integrator in its dynamics, designing a control system that guarantees robust performance is not trivial. A lane keeping support system was developed based on combining the steering torque with CCD camera measurement in [25]. Test results show that the lane keeping performance can be significantly improved, with much reduced steady-state lateral displacement, while the driver workload, measured through the frequency distributions of electromyograms of the upper arm muscles, is lowered by about half.

Despite of these positive results on combining steering torque from a human driver and the automatic control system, it is still not clear dealing with steering torque provides any advantage over steering angles. Not only is dealing with steering angle easier, in terms of vehicle dynamic equation formulation. Furthermore, combining steering angles from driver/control system is very straightforward. For example, a linear actuator to move the rack position relative to the vehicle chassis is a simple way to inject additional steering angle to the driver's steering angle.

The LKA concept has been combined with adaptive cruise control (ACC) to form a semiautomated driving system in Japan [32]. This design concept is somewhat more aggressive, in terms of market targeting, compared with the less-authoritative lane departure warning system. However, similar to the comparison between ACC and collision warning (CW) system design, CW design is, in many aspects, a more difficult engineering problem than that of ACC. This is because LKA type of systems exerts low level but continuous effort to ensure that the vehicle is operating in a safe

manner, instead of issuing a warning signal at the last moment. Therefore, if a simple and cheap actuation mechanism can be identified, LKA concept may have an easier time to gain wider acceptance compared with the lane departure warning systems.

Due to the fact that both human drivers and an automatic control system coexist, the LKA system must be designed and evaluated to ensure the system remains human centered. Driving should remain a fun task and the driver should actively participate in the driving task instead of easily falling asleep. These subtle and semisubjective requirements make it much harder to assess the performance of LKA systems.

39.2.5 Vehicle State Estimation

The bicycle model shown in Equation 39.1, even though extremely simple, very nicely represents the fundamental vehicle handling dynamics involved in the lane keeping functions. From simple observability analysis of the state-space equation, it is clear that the vehicle lateral displacement is unobservable from other measurements. In other words, lateral displacement needs to be measured, instead of being estimated. Once the lateral displacement signal is available, all the other three state variables can be estimated. Frequently, the vehicle yaw rate is also measured, which is useful for other purposes such as electronic stability program (ESP). While vehicle lateral speed can be directly measured by instrumental sensors such as optical sensors or GPS sensors, there are practical issues such as cost, accuracy, and reliability that prohibit production vehicles from using these sensors at present and in the near future. Therefore, the estimation of vehicle lateral speed based on other vehicle input/output signals is an important topic and has been widely discussed in the literature. The estimation of the other two variables (yaw rate, yaw angle error), if they are not measured, proves to be much less challenging. For example, excellent estimation of yaw rate using lateral displacement from a CCD camera is presented in Mouri et al. [33].

The problem of lateral speed estimation is equivalent to sideslip angle estimation, studied widely in the literature because of its application in ESP or four-wheel-steering applications. Senger and Kortum [34] developed a vehicle handling model and tire model to estimate the lateral speed. Their approach assumes that the tires operate in the linear region with known cornering stiffness. Kaminaga and Naito [35] applied a standard Lyapunov-based adaptive observer technique for sideslip estimation—which is equivalent to estimating vehicle lateral speed. The algorithm was found to provide satisfactory robustness for cornering stiffness variation on flat road surfaces. Experimental results provided in their paper show that the sideslip angle estimation error is below 3°. Liu and Peng [36] proposed an identification scheme for simultaneous state and parameter estimation to overcome the problems caused by unknown vehicle parameters. However, the observer stability is only proven for the case with constant cornering stiffness and/or road surfaces. It is further observed in their paper that the proposed methodology requires one to two cycles of vehicle maneuvers before the estimation converges to satisfactory performance with time-invariant cornering stiffness.

Farrelly and Wellstead [37] proposed both a physical modeling approach and a kinematics modeling approach. In their physical modeling approach, the observer can be fine-tuned to perform well even under changes of certain parameters. For example, it can be designed to be robust under cornering stiffness variation in the front axle or in the rear axle. However, the observer will not work reliably when both front and rear cornering stiffness are unknown and/or varying. As for their kinematics approach, the observer has been shown to provide satisfactory performance when the vehicle yaw rate is nonzero. Fukada [38] suggested a combined observer/direct integration method. The method strived to balance the robustness of modeling error (provided by direct integration) and that of signal bias (provided by tire model feedback). A road bank/slant estimate was also provided. In this dual approach, the lateral velocity estimate depends on the road bank estimate that, in turn, depends on the lateral velocity estimate. The experimental results on flat road show excellent performance in a double-lane change maneuver but significant error (about 40% less than the actual)

in a J-turn maneuver. The J-turn maneuver on snow is inherently difficult for this approach since the signal to noise ratio was small (as vehicle side-drifted slowly) and the tires operate in nonlinear regions. This method was subsequently refined and integrated into a vehicle stability control system [39]. The reported results suggest that the estimation algorithm was improved and performs satisfactorily.

Hac and Simpson [40] developed an algorithm that uses wheel speeds and vehicle lateral acceleration measurement, and a friction coefficient estimation scheme enhanced nonlinear vehicle model to estimate yaw rate and lateral speed simultaneously. The experimental results, especially those on icy surfaces, are impressive. However, the effect of another important factor, road bank angle, is not adequately addressed.

From the aforementioned literature review, it can be seen that the lateral speed (or sideslip angle) estimation problem has been actively studied over the last decade. A key finding, which becomes apparent in the crystal ball of hindsight, is that in order to obtain good lateral speed estimation, it is important to first obtain good measurement or estimation of the instantaneous road bank angle (superelevation). Road bank angle can be either obtained through the combined use of DGPS and lookup tables containing roadway geometric information or estimated.

In the following, the basic ideas of three different approaches for lateral speed estimation will be reviewed. These methods include a transfer function approach, the state-space approach by Liu and Peng [36], and the kinematics approach by Farrelly and Wellstead [37]. In addition, a method for dynamic road bank angle estimation [41] will also be reviewed.

39.2.5.1 Kalman Filter Method

The concept of this estimation method is again based on the bicycle vehicle model shown in Equation 39.1. However, to include the effect of an important disturbance input—the road bank angle into the dynamic equation—it is modified slightly:

$$\frac{d}{dt}\begin{bmatrix} y \\ v \\ \psi-\psi_d \\ r \end{bmatrix} = \begin{bmatrix} 0 & 1 & u & 0 \\ 0 & -\dfrac{C_{\alpha f}+C_{\alpha r}}{mu} & 0 & \dfrac{bC_{\alpha r}-aC_{\alpha f}}{mu}-u \\ 0 & 0 & 0 & 1 \\ 0 & \dfrac{bC_{\alpha r}-aC_{\alpha f}}{I_z u} & 0 & -\dfrac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z u} \end{bmatrix}\begin{bmatrix} y \\ v \\ \psi-\psi_d \\ r \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{C_{\alpha f}}{m} \\ 0 \\ \dfrac{aC_{\alpha f}}{I_z} \end{bmatrix}\delta + \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} r_d + \begin{bmatrix} 0 \\ g \\ 0 \\ 0 \end{bmatrix}\gamma \tag{39.2}$$

where γ is road superelevation angle in radian. The small-angle assumption is again used so that this equation remains linear. By assuming that the vehicle is roughly aligned with the tangential direction of the roadway, the gravity force, acting through the small road bank angle, directly influences the vehicle lateral motion but not the yaw motion. The standard Kalman filter technique can be applied to Equation 39.2 to obtain the lateral speed estimation. In this problem formulation, both the steering angle and road curvature were assumed to be known inputs. The road bank angle, on the other hand, was assumed to be a Gaussian and white disturbance. Commonly, two vehicle motion sensors are used to provide crucial measurements for lateral speed estimation—vehicle lateral displacement (y) and yaw rate (r). A direct application of this standard technique using Equation 39.2, however, was found to produce unsatisfactory results. There are two major reasons. First, as stated before, the road bank angle directly influences the vehicle lateral motion, and, thus, without an accurate representation of this important variable, it is very difficult to obtain good estimation of the lateral vehicle speed. Secondly, the vehicle yaw rate measurement may contain unknown offset arising from alignment, fluctuation of vehicle electrical potential, etc. These two problems can be dealt with by augmenting the original bicycle model with two extra state variables [42], namely,

$$\frac{d}{dt}\begin{bmatrix} y \\ v \\ \psi-\psi_d \\ r \\ r_{bias} \\ \gamma \end{bmatrix} = \begin{bmatrix} 0 & 1 & u & 0 & 0 & 0 \\ 0 & -\frac{C_{\alpha f}+C_{\alpha r}}{mu} & 0 & \frac{bC_{\alpha r}-aC_{\alpha f}}{mu}-u & 0 & g \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{bC_{\alpha r}-aC_{\alpha f}}{I_z u} & 0 & -\frac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z u} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ v \\ \psi-\psi_d \\ r \\ r_{bias} \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{C_{\alpha f}}{m} & 0 \\ 0 & -1 \\ \frac{aC_{\alpha f}}{I_z} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta \\ r_d \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 \\ \frac{C_{\alpha f}}{m} & 0 & 0 \\ 0 & 0 & 0 \\ \frac{aC_{\alpha f}}{I_z} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \delta_{dist} \\ \dot{\gamma} \\ \dot{r}_{bias} \end{bmatrix} \tag{39.3}$$

In Equation 39.3, the measured steering angle δ and road curvature are grouped as the known input vector. Three unknown disturbance inputs are assumed to exist: an unknown steering term, road bank angle rate, and yaw rate bias rate. The augmented system was found to provide improved estimation mainly due to a more accurate representation of the two augmented state variables [42].

39.2.5.2 Transfer Function Method

The vehicle lateral speed can be estimated without using the four-state bicycle model. Instead, the more traditional two-state system can be used. The two-state system, of course, implies that we are not aware of (or do not care about) the road following performance of the vehicle. In addition, if (a big if!) the road bank angle is negligible, the two-state bicycle model can be written as

$$\begin{bmatrix} \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} -\frac{C_{\alpha f}+C_{\alpha r}}{mu} & \frac{bC_{\alpha r}-aC_{\alpha f}}{mu}-u \\ \frac{bC_{\alpha r}-aC_{\alpha f}}{I_z u} & -\frac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z u} \end{bmatrix} \begin{bmatrix} v \\ r \end{bmatrix} + \begin{bmatrix} \frac{C_f}{m} \\ \frac{aC_f}{I_z} \end{bmatrix} \delta \tag{39.4}$$

If we rewrite Equation 39.4 in a simpler form,

$$\begin{bmatrix} \dot{v} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix} \begin{bmatrix} v \\ r \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \delta \tag{39.5}$$

then it is straightforward to write the transfer functions from steering angle δ to the two states as

$$H_{\delta \to v} = \frac{sb_1 - a_4 b_1 + a_3 b_2}{s^2 - s(a_1 + a_4) + a_1 a_4 - a_2 a_3} \tag{39.6}$$

and

$$H_{\delta \to r} = \frac{sb_2 + a_2 b_1 - a_1 b_2}{s^2 - s(a_1 + a_4) + a_1 a_4 - a_2 a_3} \tag{39.7}$$

The basic idea of the transfer function method is as follows: we can first use the available measurements (steering angle and yaw rate) to obtain a least squares fit for the transfer function shown in Equation 39.7 in real time. This optimal ARMA model (see Chapter 6), which best describes the vehicle yaw motion in a discrete-time format of Equation 39.7, will then be converted into the continuous time where best-guess parameters of the transfer function for the yaw rate output (Equation 39.7) can be obtained. Since the parameter b_2, which is the coefficient of the "s" term of the numerator polynomial of Equation 39.7, is related to the front tire cornering stiffness in a simple form ($b_2 = aC_f / I_z$, see Equations 39.4 and 39.5), it is straightforward to calculate a best-guess front axle cornering stiffness, C_f, the value of which can be used to compute $b_1 = C_f/m$. These values can then be plugged into the constant term of the numerator polynomial of Equation 39.7, $a_2 b_1 - a_1 b_2$, to calculate the rear axle cornering stiffness C_r. Finally, using the computed axle cornering stiffness values, the numerator polynomial of the lateral speed transfer function shown in Equation 39.6 can be calculated. Since the two transfer functions share the same denominator, the transfer function in Equation 39.6 is completely known, based on which the vehicle lateral speed can be estimated.

Since the discrete-time ARMA identification process requires only measured variables in three history steps, the adaptation for tire–road characteristics change is very fast. It is important to point out that the basic idea of this approach is simple. In practice, it was found that the identified least squares model may become unstable or very lightly damped, especially when the vehicle yaw rate rises quickly, responding to fast steering input. Under these circumstances, a commonly adopted practice is to discard these unstable models and just keep using the non updated (but stable and well-damped) model. This modification is necessary to keep the estimated tire cornering stiffness positive and to ensure that the model prediction stays within the reasonable region. The selection of the threshold value was found to have significant effect on the final estimation results [43].

39.2.5.3 State-Space Method

The idea presented in Section 39.2.5.2 can be extended to obtain best estimate of a state-space method, based on which vehicle lateral speed can be estimated. However, this two-step approach heavily relies on the structure of the transfer function and is frequently sensitive to uncertainties, such as variation in other vehicle parameters (m, I_z, etc.). In the following, the adaptive algorithm developed in [36] is described. This state-space algorithm was developed to estimate states and unknown parameters simultaneously for nonlinear time-invariant systems that depend affinely on the unknown parameters. The lateral speed estimation problem turns out to have the required structure, due to the fact that the tire cornering stiffness of the two axles is generally unknown and is the main source of model uncertainties. It is thus possible to estimate vehicle lateral speed and the affine road friction parameter simultaneously. Equation 39.4 can be rewritten in the linearly parameterized form as

$$\dot{x} = Ax + u_c \theta \tag{39.8}$$

$$y = r$$

where

$$x = \begin{bmatrix} v \\ r \end{bmatrix}$$

$$A \equiv \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix}$$

u_c is the front axle steering angle

$$\theta = \begin{bmatrix} \frac{2C_f}{m} & \frac{2aC_f}{I_z} \end{bmatrix}^T$$

The adaptive observer is then

$$\dot{\hat{x}} = (A_n + \Delta\hat{A}_x)\hat{x} + \Delta\hat{A}_y r + \begin{bmatrix} -u \\ 0 \end{bmatrix} r + u_c\hat{\theta} + K(y - \hat{y}) \tag{39.9}$$

where

$$\hat{x} = \begin{bmatrix} \hat{v} & \hat{r} \end{bmatrix}^T$$

$$\hat{y} = \hat{r}$$

$$A_n = \begin{bmatrix} -\left(\frac{2C_{fn} + 2C_{rn}}{um_n}\right) & \left(\frac{-2a_nC_{fn} + 2b_nC_{rn}}{um_n}\right) \\ \left(\frac{-2a_nC_{fn} + 2b_nC_{rn}}{uI_{zn}}\right) & \left(\frac{-2a_n^2C_{fn} - 2b_n^2C_{rn}}{uI_{zn}}\right) \end{bmatrix}$$

$$\Delta\hat{A}_x = \begin{bmatrix} \Delta\hat{a}_1 & 0 \\ \Delta\hat{a}_2 & 0 \end{bmatrix}$$

$$\Delta\hat{A}_y = \begin{bmatrix} \Delta\hat{a}_3 & \Delta\hat{a}_4 \end{bmatrix}^T$$

The subscript "n" is used to denote the nominal parameters. For example, C_{fn} is the nominal value of C_f. The unknown variables $\Delta\hat{a}_i$ denote the deviation of a_i from their nominal values (entries in the A_n matrix), which are to be estimated. If the observer gain is selected as $K = [k_1, k_2]^T / u$, that is,

$$A_n - KC = \frac{1}{u}\begin{bmatrix} -\left(\frac{2C_{fn} + 2C_{rn}}{m_n}\right) & \left(\frac{-2a_nC_{fn} + 2b_nC_{rn}}{m_n}\right) - k_1 \\ \left(\frac{-2a_nC_{fn} + 2b_nC_{rn}}{I_{zn}}\right) & \left(\frac{-2a_n^2C_{fn} - 2b_n^2C_{rn}}{I_{zn}}\right) - k_2 \end{bmatrix} \tag{39.10}$$

where C = [0 1], then the closed-loop state matrix is stable when k_1 and k_2 are properly selected, according to the nominal vehicle parameters. When ΔA_x, ΔA_y, and θ are constant (slow varying), these unknown parameters, as well as the state variables, are estimated by using a standard state

observer (which uses estimated parameters for its model), together with a parameter update law, which is gain varied based on the calculation of a set of augmented states. The equations are too numerous to recite completely in this chapter. Interested readers are referred to the original paper by Liu and Peng for details. This estimator, like most other adaptive algorithms, requires persistent excitation condition to achieve small estimation errors. In other words, when the vehicle is driving straight, the vehicle parameters will not be updated accurately. In other words, if the transition from high-friction to low-friction road surfaces occurs during a straight section of the roadway, this algorithm will need some time to update its estimation and thus might suffer slow convergence transiently. The most important design parameters of this estimator were found to be the observer gains k_1 and k_2, which need to be tuned properly according to the trade-off between noise rejection and convergence speed. We also found it necessary to constrain the estimated parameter values, to ensure that they stay within reasonable values governed by the physical parameters.

39.2.5.4 Kinematics Method

The kinematics approach proposed by Farrelly and Wellstead [37] is promising as the stability of the observer/convergence of estimation error has been shown for nonzero yaw rate. In the following, their methodology is integrated with a physical model-based observer to avoid unobservability and the drifting issues (as described in [37]) during near-zero yaw rate conditions. The implemented observer arbitrates between a kinematics model-based observer and a dynamical model-based observer depending on the magnitude of the yaw rate—the dynamical model will be used only when the vehicle yaw rate is large enough. The model is described in the following:

If ($|r| > r_c$ deg/s),

$$\begin{bmatrix} \dot{\hat{u}} \\ \dot{\hat{v}} \end{bmatrix} = \begin{bmatrix} 0 & r \\ -r & 0 \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix} + \begin{bmatrix} a_x \\ a_y \end{bmatrix} + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} (\hat{u} - u) \tag{39.11}$$

where

$$k_1 = 2|\alpha|r,\ k_2 = (\alpha^2 - 1)r,$$

else

$$\dot{\hat{v}} = -ur + \hat{a}_y,\ \hat{u} = u$$

where

$$\hat{a}_y = \frac{[2C_f(\tan^{-1}((\hat{v} + ar)/u) - \delta) + 2C_r \tan^{-1}((\hat{v} - br)/u)]}{m}$$

Note that r_c and α are design parameters and δ, *u, r,* a_x, and a_y are measured variables. Since the sensor measurements of current generation of vehicle stability controller usually include δ, *u, r,* and a_y but not a_x, we further adapt this method to accommodate current sensor platform by approximating a_x with vehicle speed information, that is, $a_x = du/dt$. While this approximation adds sensor measurement noise/error in the longitudinal acceleration signal, a_x, used in the observer, there are also other real-world sensor noises in vehicle speed and longitudinal and lateral acceleration measurements due to wheel slip, sensor drift, cross talk induced by vehicle motions, and road noise. In the following, we analyze the estimation performance of the modified F&W approach (henceforth

referred as the kinematics approach) under these sensor noises, that is, u, a_x, and a_y. Equation 39.11 can be rewritten in the following form:

$$\begin{bmatrix}\dot{\tilde{u}}\\ \dot{\tilde{v}}\end{bmatrix}=\begin{bmatrix}-2\alpha|r| & r\\ -\alpha r & 0\end{bmatrix}\begin{bmatrix}\tilde{u}\\ \tilde{v}\end{bmatrix}+\begin{bmatrix}a_{xn}\\ a_{yn}\end{bmatrix}+\begin{bmatrix}-2\alpha|r|\\ -\alpha r\end{bmatrix}u_n \tag{39.12}$$

where the variables with ~ denote estimation error. If we use a Lyapunov function $V=(\alpha^2\tilde{u}^2+\tilde{v}^2)/2$, it can be seen that

$$\dot{V}=-2\alpha^3|r|\tilde{u}^2+\alpha^2\tilde{u}\cdot a_{xn}+\tilde{v}\cdot a_{yn}+[2\alpha^3|r|\tilde{u}+(\alpha^2-1)r\tilde{v}]\cdot u_n$$

Unlike the original analysis in [37], *dV/dt* is no longer negative definite when sensor uncertainties exist, and the design parameter α influences the significance of sensor measurement noise. Furthermore, from Equation 39.12, we have $G_{a_{xn}\to\tilde{v}}=-\alpha^2 r/(s+\alpha|r|s)$, $G_{a_{yn}\to\tilde{v}}=s+2\alpha|r|/(s+\alpha|r|)^2$, and $G_{v_{xn}\to\tilde{v}}=(\alpha^2-1)rs-2\alpha r|r|/(s+\alpha|r|)^2$. From these equations, it is obvious that the design parameter α influences the error propagation speed of longitudinal acceleration noise, the sensitivity to lateral acceleration noise, and the sensitivity to vehicle speed noise. It can also be seen that the steady-state error contributed from our approximation of longitudinal acceleration signal is $-v$ since the approximation error we introduced is in the amount of vr and the steady-state gain is $1/r$. In other words, the estimated lateral speed will eventually be washed out to zero if the derivative of vehicle speed is used in place of longitudinal acceleration. On the other hand, the estimation error during transient events (which are most critical for real-world vehicle stability control) is negligible if we keep the observer gain α small.

With real-world environment noise, a trade-off decision has to be made for the observer gain α since increasing the observer gain increases the robustness to measurement noise of a_y and u but decreases the robustness (of estimation during transient events) to measurement noise of a_x.

39.2.5.5 Road Bank Angle Estimation

As stated earlier, the road bank angle estimation problem is related to the lateral speed estimation problem. They use the same (similar) set of vehicle lateral dynamic equations, and an accurate road bank angle information will dramatically help the lateral speed estimation. This is true for estimation problems based on the four-state bicycle model [42] and the two-state bicycle model [43]. There is a chicken-and-egg problem here: Road bank angle estimation problem is frequently solved by assuming certain lateral speed condition (e.g., $\dot{v}\approx 0$). The estimated road bank angle is then used to estimate lateral speed, to verify whether the assumption is correct or not. Alternatively, the road bank angle can be estimated based on concepts similar to the transfer function approach described in 39.2.5.2, which are less restrictive but in the meantime are also more sensitive to measurement noise and transient response—transfer function is fundamentally a steady-state response method, so its performance during transient is not guaranteed. An approach to combine the dynamic equation (assuming $\dot{v}\approx 0$) and transfer function method is described in the succeeding text, which is nothing but an excerpt of the work shown in [41].

Expanding Equation 39.4 to include road bank angle effect, we have

$$\begin{bmatrix}\dot{v}\\ \dot{r}\end{bmatrix}=\begin{bmatrix}-\dfrac{C_{\alpha f}+C_{\alpha r}}{mu} & \dfrac{bC_{\alpha r}-aC_{\alpha f}}{mu}-u\\ \dfrac{bC_{\alpha r}-aC_{\alpha f}}{I_z u} & -\dfrac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z u}\end{bmatrix}\begin{bmatrix}v\\ r\end{bmatrix}+\begin{bmatrix}\dfrac{C_f}{m}\\ \dfrac{aC_f}{I_z}\end{bmatrix}\delta+\begin{bmatrix}g\\ 0\end{bmatrix}\sin(\phi)$$

$$\equiv\begin{bmatrix}a_1 & a_2\\ a_3 & a_4\end{bmatrix}\begin{bmatrix}v\\ r\end{bmatrix}+\begin{bmatrix}b_1\\ b_2\end{bmatrix}\delta+\begin{bmatrix}g\\ 0\end{bmatrix}\sin(\phi) \tag{39.13}$$

based on which the transfer function from steering angle and road bank excitation (*sin*(φ)) to vehicle yaw rate and lateral acceleration can be found:

$$G_{\delta \to r} = \frac{b_2(s-a_1)+a_3b_1}{s^2-(a_1+a_4)s+a_1a_4-a_2a_3} \tag{39.14}$$

$$G_{\delta \to a_y} = \frac{s\left[b_1(s-a_4)+a_2b_2\right]+u\left[b_2(s-a_1)+a_3b_1\right]}{s^2-(a_1+a_4)s+a_1a_4-a_2a_3} \tag{39.15}$$

$$G_{\sin(\phi)\to r} = \frac{a_3 g}{s^2-(a_1+a_4)s+a_1a_4-a_2a_3} \tag{39.16}$$

$$G_{\sin(\phi)\to a_y} = \frac{sg(s-a_4)+ua_3g}{s^2-(a_1+a_4)s+a_1a_4-a_2a_3} \tag{39.17}$$

where u is the vehicle forward speed, and the relationship $a_y = \dot{v} + ur$ is used. Given these transfer functions, the measured vehicle yaw rate and lateral acceleration signals can be written as

$$r_{measured} = r_{actual} + r_{noise} \approx G_{\delta \to r} \cdot \delta + G_{\sin(\phi)\to r} \cdot \sin(\phi) + r_{noise} \tag{39.18}$$

$$a_{y_{measured}} = a_{y_{actual}} + a_{y_{noise}} \approx G_{\delta \to a_y} \cdot \delta + G_{\sin(\phi)\to a_y} \cdot \sin(\phi) + a_{y_{noise}} \tag{39.19}$$

Notice that in Equations 39.18 and 39.19, the approximate sign "≈" was used because of the fact that transfer functions do not capture transient response and that the actual vehicle response might be nonlinear. Based on Equations 39.18 and 39.19 and assuming that the sensor noise is Gaussian white and thus cannot be easily dealt with, two road bank estimations can be obtained:

$$\sin(\hat{\phi}_r) = G_{\sin(\phi)\to r}^{-1}\left(r_{measured} - G_{\delta \to r} \cdot \delta\right) \tag{39.20}$$

$$\sin(\hat{\phi}_a) = G_{\sin(\phi)\to a_y}^{-1}\left(a_{y_{measured}} - G_{\delta \to a_y} \cdot \delta\right) \tag{39.21}$$

It should be noted that the inverse of the transfer functions $G_{\sin(\phi)\to r}^{-1}$ and $G_{\sin(\phi)\to a_y}^{-1}$ is noncausal. In reality, these noncausal continuous-time transfer functions can be approximated by causal discrete-time operators, or steady-state gains. In [41], a third road bank angle estimation is suggested, which is based on the assumption $\dot{v} \approx 0$:

$$\sin(\hat{\phi}_v) = \frac{\left(u \cdot r_{measured} - a_{y_{measured}}\right)}{g} \tag{39.22}$$

Since all three estimations are based on certain simplification assumptions, they are accurate only within certain operational conditions. By figuring out the current situations, these three estimates can be selected or combined to obtain the best estimation.

39.3 PRECISE LATERAL MANEUVER

While low-speed operations can be challenging for longitudinal controls due to the nonlinearities of power train, tire, and braking systems, low vehicle forward speed poses very different challenges to lateral controls. At higher speeds, the steering angle becomes a more potent input to influence vehicle lateral motions. The lateral control algorithm designed for operating under high vehicle forward speeds thus needs to accurately model the tire forces and vehicle rigid-body dynamics. Due to the higher level of lateral accelerations, the coupling between vehicle roll motion and lateral/yaw motions becomes non negligible.

Some researchers argued that the bicycle model, used extensively in this chapter, fails to capture the frequency response of vehicle lateral acceleration around 1–2 Hz [44], especially when the vehicle has a soft suspension, a high CG (e.g., a bus), or both. A lateral controller to achieve precise tracking performance, or robust operations, thus may need to be designed based on a 3-DOF (lateral, yaw, and roll) model, instead of the bicycle model [44,45]. Since this argument is supported by experimental data, it seems to be hard to argue against. However, the author wishes to provide an alternative view in the following.

If the main purpose of the vehicle model is to describe the vehicle lateral motion, relative to the roadway (which is the case for automatic lane following), then the vehicle lateral acceleration, measured by an accelerometer installed on the sprung mass, is a poor measurement. This is because the measured signal is contaminated by the sprung mass roll motion and thus does not reflect the lateral motion that we truly care about. In [44], experimental comparison of both yaw rate and lateral acceleration is provided. If we focus on the yaw rate comparison, it can be seen that the 2-DOF (bicycle) model and the 3-DOF model are almost identical up to 2 Hz. At higher frequencies, the linear 3-DOF model may not be trustworthy, because it did not capture load transfer and associated nonlinear tire dynamics. Therefore, the value of using the linear 3-DOF model for the design of lateral control algorithm is limited. Its true value perhaps lies in other applications. For example, for modern vehicle without an underbody frame, it is very hard, if not impossible, to install an accelerometer "on the unsprung mass." The 3-DOF model can be used for signal processing purposes to provide the lateral acceleration of the "unsprung mass," load transfer, etc.

At low forward speeds, a very different problem arises—the vehicle lateral dynamics quickly become uncontrollable as the vehicle forward speed decreases. This can be understood by examining the extreme case—when the vehicle forward speed is zero, steering angle loses its effectiveness in generating lateral forces—and thus the ability to manipulate vehicle lateral motions. Therefore, at a low forward speed, large steering angle needs to be used, which introduces a new sort of problems—nonlinearities due to the steering and suspension system, tire mechanics, and frictions. The duo problem—the vehicle also becomes unobservable as forward speed becomes zero—is less intuitive. There are two ways this can be understood: First, frequently we rely on vehicle lateral displacement (1st state of Equation 39.1) to "observe" the lateral speed and/or yaw angle error (2nd and 3rd state of Equation 39.1). When vehicle speed is zero, a car is parked, and, thus, the same lateral displacement could correspond to infinitely many yaw angle error values. Secondly, since the vehicle forward speed appears at the denominator of the state matrix, zero forward speed corresponds to singularity of several matrix elements. The second way to look at the vehicle observability relies more on mathematics and is more rigorous.

At low forward speeds, a vehicle becomes "less controllable" and "less observable" and thus both control and state estimation may become much more challenging. It is a good idea to stop using modern full-state feedback algorithms and revert back to simple, output (lateral displacement) feedback control algorithm at very low forward speed.

Due to the reduction of dynamic forces at lower forward speeds, the vehicle lateral control problem changes from dynamic to a kinematic one and becomes an easier problem from that aspect. So why are we choosing to have a separate section on "precise maneuver," instead of devoting the

section to other areas with more abundant literature (e.g., lane change [1–3])? The reason is that this is an area that imposes its own technical challenges. Lane change, as we have argued before, is mainly a decision problem involving several vehicles in multiple lanes. The decision is more dependent on vehicle longitudinal positions than lateral motions. If the coordination with vehicle longitudinal motion is not considered (e.g., assuming some other systems were trusted with that decision), lane change is nothing but a simple extension of lane following technique, involving trajectory planning for smooth operations or special methods to address the sensing issues. Therefore, we do not think it warrants another section to discuss about lane change modeling and control.

Two low-speed applications that are likely to see adoption in the near term are urban bus automations and automated valet systems. These two concepts are described in the following subsections.

39.3.1 Automated Buses

Automated buses are concepts competing with existing alternatives such as light rails and human-controlled buses. Through automation, buses can mimic the operation of light rails and achieve very precise maneuvers. Precise tracking performance is important when the cheaper and less-intrusive "unprotected configurations" are used, that is, no fence/barricade will be installed to separate the buses with other vehicles/pedestrians. When the (either protected or unprotected) lanes are instrumented and dedicated to be used by automated buses or in some cases coexist with manually driven buses, a bus rapid transit (BRT) system with reliability and capacity approaching that of rail transit systems becomes possible. This new transit concept promises much reduced capital investment, increased routing and rerouting flexibility, and much lower operation and maintenance costs.

Automated buses working based on magnetic markers were found to achieve a "docking" accuracy of less than 1 cm. This level of accuracy is much better than that achievable by human drivers and offers improved safety in an unprotected environment. The accuracy is obtained at slow speeds, where vehicle controllability and observability are low.

39.3.2 Parking Assist and Automated Valet

Two closely related concepts are discussed in this section. Parking assist addresses a task that is among the most dreadful to novice drivers. Parallel parking, involving back-and-forth maneuvering, sometimes in a very tight spot, requires good feel of the vehicle position and precise timing of the reversal of large steering action. Sometimes the available parking space is just a little wider and/or longer than the vehicle, and, thus, multiple back-and-forth maneuvers need to be planned and executed sequentially.

One of the earliest parking assist system concepts was developed by Aisin Seiki [46,47] and implemented on the Japanese version of Toyota Prius model introduced in September 2003. This system utilizes an electrical power steering system and backup guide monitor technology and can be used for parallel or garage parking. The system hardware consists of a rearview camera, steering angle sensor, vehicle speed sensor, display, speakers, yaw rate sensor, and an electronic control unit.

In both the "normal" mode (garage parking) and the "parallel parking" mode, the rearview camera calculates and projects the vehicle motion and overlays important features (e.g., rear bumper, "parking frame") or guidance curves onto the display. The basic idea is that the projected vehicle trajectory can be used by the human driver, who controls the steering to move the projected vehicle trajectory into the parking space. In the parallel parking case, the operation is somewhat more complicated because a reverse in steering action is necessary. The details are omitted here and interested readers can refer to [46].

The basic concept of automated valet systems, which are not yet available commercially, is quite different from parking assist systems. As the name implies, a human driver will not be present and the valet system is expected to drive the vehicle automatically and safely by itself. Both vehicle speed and steering will be controlled and mostly under very low speed. Automated valet is a natural

extension of the parking assist system but is expected to be confined in a small niche market. To replace the human driver, it is necessary to ensure that the operation is under one of these two conditions: (1) either enough intelligence and robustness can be included in the system, through accurate obstacle/environment detection, classification, and control, or (2) when the required level of intelligence/robustness is not available, ensure that both the parking facility and the vehicles are highly structured and heavily instrumented. Option (1) is much more challenging but in the meantime has much larger potential in the near term. Option (2) already exists in certain primitive form as automatically guided robots. The need for customized infrastructure and highly instrumented vehicles will severely limit its applications.

One possible design concept for automatic valet system will rely on centralized sensing systems to reduce the need of expensive on-vehicle equipment. For example, an overhead camera installed to watch over the parking space can provide information for individual vehicle position as well as the environment, including available spaces, vehicles moving in a blind spot, or oncoming in an intersection. When drive-by-wire becomes widely available, this centralized sensing plus decentralized control architecture will make automatic valet systems a reality.

39.4 LANE DEPARTURE WARNING AND PREVENTION

Lane departures are important vehicle handling problems because they frequently result in serious accidents, including run-off-road crashes, head-on collision with oncoming vehicles, or collision with highway divider guard rails or posts. Road departure crashes are one of the leading causes of highway fatalities in the United States. The USDOT defines an "off-roadway crash" as one in which the first harmful event occurs off the roadway. An estimated 6.1 million police-reported crashes involving light vehicles occurred in the United States in 2000, causing approximately 41,000 fatalities [48]. Of these crashes, the USDOT estimates that approximately 1.12 million are associated with a single vehicle departing the roadway due to inadvertent lateral drifting or loss of control and account for about 41% of all fatalities. Thus, road departure accidents not only form a significant portion of all police-reported accidents but also involve more fatalities than other accidents.

Lane departures usually arise from driver inattention, relinquished steering (due to physiological reasons such as falling asleep, drug or alcohol, or seizure), or loss of directional control due to adverse road condition, high vehicle speed, or sudden evasive maneuvers. An effective lane departure warning or prevention system, therefore, will necessarily play the role of a "backseat driver" that constantly watches the vehicle handling motion and issues warnings or intervenes when a lane departure is determined to be imminent.

The basic function of the lane departure warning/prevention system is to project vehicle trajectory and compare with the perceived road geometry (obtained from vision systems) to calculate a performance metric. A classic metric for determining impending road departure is time-to-lane crossing (TLC), which can be understood as the time until the vehicle CG will cross either edge of the roadway under certain conditions [49]. This definition of TLC in its truest sense usually results in an extremely large value of TLC for most driving conditions. This is because the vehicle is controlled by a human driver who is, most of the time, doing a good job. If the driver relinquishes steering so that the vehicle drives an open loop (which can be easily experimented on an open highway), due to various disturbances (road crown, wheel alignment, etc.), TLC is generally around 8–12 s. If this (unsafe) experiment is done with a passenger on board, he or she likely will speak out when TLC approaches 3–4 s. Based on these observations, human-centered lane departure warning/prevention systems can be designed to keep track of TLC, for values up to about 4–5 s, with the warning/intervention threshold set at about 1–2 s. The statistical distribution of TLC can be obtained by keeping track of TLC histogram for typical human drivers; one such histogram can be found in [50]. The calculation of TLC will be discussed in Section 3.1

To project the vehicle future trajectory accurately (so that TLC can be computed), we need to update vehicle parameters (speed, cornering stiffness, etc.) and estimate dominant external

disturbances (road superelevation). The estimation of external disturbances, however, also depends on the vehicle parameters. Therefore, on-line estimation of road/tire characteristics and road superelevation is crucial for the TLC calculation and the overall lane departure warning system. Since the road superelevation (a.k.a. bank angle) estimation has been discussed in Section 39.2.5.5, we will only focus on the estimation of tire–road friction or cornering stiffness in Section 39.4.1.

39.4.1 Calculation of Time-to-Lane Crossing

Here, in this section, we will assume that the problem at hand is as follows: Given all the vehicle/road parameters (mass, inertia, CG location, forward speed, cornering stiffness, road bank angle, and future road curvature) and the available onboard measured signals (steering angle, lateral displacement, yaw rate, and lateral acceleration), calculate TLC accurately.

As stated before, TLC is defined as the time until the vehicle CG is projected to cross either edge of the roadway. There are two ways TLC can be calculated. A simpler way is to calculate TLC from [42]

$$\begin{aligned} \dot{y} &= u \cdot (\psi - \psi_d) + v \\ \text{TLC} &\equiv \frac{y'}{\dot{y}} \quad \dot{y} \neq 0 \end{aligned} \tag{39.23}$$

where

y is the vehicle lateral displacement from the center of the road
u is the vehicle forward speed
$\psi - \psi_d$ is the relative heading angle
v is the vehicle lateral speed

y' is defined as $y \pm$ *lane width*/2, and depending on the sign of $\dot{y}$, only the one that is reducing to zero is used in the calculations. This calculation is based on linear extrapolation, that is, assuming the rate of change of the lateral displacement $\dot{y}$ remains constant into the foreseeable future. This assumption is reasonable when the vehicle is slowly drifting away on a straight roadway, but will not be accurate when large road curvature or steering is involved. Obviously, to implement Equation 39.23, two vehicle state variables need to be estimated: $\psi - \psi_d$ and v. These estimations can be obtained from the vehicle state estimation algorithms discussed in Section 39.2.5 and thus will not be repeated here.

The second, popular way to calculate TLC is to use a vehicle model and keep integrating the model into the future [51]. For example, based on Equation 39.3, a model with no unknown disturbances can be constructed:

$$\frac{d}{dt}\begin{bmatrix} y \\ v \\ \psi - \psi_d \\ r \\ r_{bias} \\ \gamma \end{bmatrix} = \begin{bmatrix} 0 & 1 & u & 0 & 0 & 0 \\ 0 & -\dfrac{C_{\alpha f} + C_{\alpha r}}{mu} & 0 & \dfrac{bC_{\alpha r} - aC_{\alpha f}}{mu} - u & 0 & g \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \dfrac{bC_{\alpha r} - aC_{\alpha f}}{I_z u} & 0 & -\dfrac{a^2 C_{\alpha f} + b^2 C_{\alpha r}}{I_z u} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ v \\ \psi - \psi_d \\ r \\ r_{bias} \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \dfrac{C_{\alpha f}}{m} & 0 \\ 0 & -1 \\ \dfrac{aC_{\alpha f}}{I_z} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta \\ r_d \end{bmatrix} \tag{39.24}$$

If the future road curvature is known (from the forward-looking sensor) and the steering angle is assumed to be fixed, both of the two input signals (δ and *rd*) are known. Therefore, Equation 39.24 can be integrated into the future until the lateral displacement is found to cross either of the two lane edges. The time at this lane crossing event is then the TLC. Under normal conditions, TLC may be very large, so this forward simulation can be terminated at, say, 4–5 s. Beyond this value, the situation is safe and it is not necessary to pinpoint the actual value of TLC.

When one implements a lane departure warning/prevention system, TLC needs to be calculated within the control sampling time, for example, 0.1 s. So direct integration of Equation 39.24 may not be possible on embedded automotive ECUs. The computation burden can be greatly reduced by calculating state transition matrices of linear dynamic systems off-line, namely, for a linear system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{39.25}$$

If the system has initial condition x_0 and a constant input over the horizon $[0,t]$ is applied to the system, then $x(t) = e^{At}x_o + \int_0^t e^{A\eta}Bud\eta$ and $y(t) = Cx = Ce^{At}x_o + C\left[\int_0^t e^{A\eta}d\eta\right]Bu$. By defining $F(t) = Ce^{At}$ and $g(t) = C\left[\int_0^t e^{A\eta}d\eta\right]B$, these matrices can be computed off-line at predetermined intervals, for example, every 0.1 s; then the output variable $y(t)$ can be calculated quickly from simple multiplications and additions from $y(t) = F(t)x_o + g(t)u$. This process can be repeated for $t = [0.1, 0.2, ...t_f]$, where t_f is the upper bound for TLC calculation.

39.4.2 Estimation of Tire Cornering Stiffness

In the previous section, TLC was calculated by assuming the state matrix and input matrix of the linear system is known. In reality, some of the vehicle parameters may change, and, among them, the tire cornering stiffness is perhaps the most difficult to estimate. Cornering stiffness is a lumped parameter that could be influenced by tire normal load (thus road roughness and vehicle load), tire pressure, road material and surface condition, temperature, etc. One estimation scheme, which assumes that the 2-DOF bicycle model is a good approximation of the vehicle handling dynamics, was introduced in Section 39.2.5.2. This method uses the vehicle lateral/handling response for cornering stiffness estimation, thus providing slower estimation. A recent research uses DGPS and lateral dynamics for cornering stiffness estimation report convergence in about 5–10 s [52]. This level of response rate may not be fast enough for lane departure applications.

In the literature, most of the tire–road friction estimation schemes use features of tire behavior (e.g., wheel speed, wheel acceleration, aligning moment, tire noise) as the basis of the estimation. For example, Eichhorn and Roth [53] used optical and noise sensors at the front end of the tire and stress and strain sensors inside the tire's tread and studied both "parameter-based" and "effect-based" road friction estimation methods. Ito et al. [54] used the applied traction force and the resulting wheel speed difference between driven and non driven wheels to estimate the road surface condition. Pal et al. [55] applied the neural network-based identification technique to predict the road frictional coefficient based on steady-state vehicle response signals. Pasterkamp and Pacejka [56] developed an on-line estimation method based on lateral force and self-aligning torque measurements. Gustafsson [57] developed a model-based approach and used the difference between driven and undriven wheels to detect tire–road friction. Liu and Peng [58] applied the special structure of the brush tire model and used wheel speed signal to estimate road friction coefficient and cornering stiffness.

The paper by Umeno et al. [59,60] is based on a very different concept. They observed that the wheel speed measurement actually contains a vibration component around 40 Hz, which arises from torsional compliance at the wheel hub. When the power spectrum of the wheel speed measurement is plotted, a peak around 40 Hz is very obvious when the tire drives on dry asphalt. The magnitude of this resonance is much reduced when the road surface is icy. Therefore, by detecting the strength of the resonance at 40 Hz, road friction can be detected. An extremely important benefit of this method lies in the fact that the vehicle can be cruising with very small longitudinal slip, yet the torsional vibration and its resonance can still exist. Therefore, this detection algorithm works without the need of large longitudinal excitations. A drawback of this method is its heavy computational requirement. The same vibration phenomenon has also been successfully applied to detect change of tire pressure [61], another important phenomenon that influences tire cornering stiffness.

After the tire–road friction coefficient is computed, cornering stiffness can be updated and TLC can be more accurately computed.

39.4.3 Warning and Control Algorithms

One reason the TLC is a popular metric for warning/control action is the fact that it clearly indicates the threat level, measured in time [62]. In addition, the threat is detected before a danger occurs. This is a big improvement over the current practice—roadside rumble strips (which give warning after a lane departure has occurred). Recently, a future offset distance (FOD) warning algorithm is proposed [63], which can be understood as constructing a virtual lane boundary as the basis for warning. When the vehicle is projected to cross the virtual lane boundary, warning will be triggered. This can be understood as a virtual (or variable) rumbling strips (VRS) concept [64].

When the threat metric is selected, the next question is as follows: What is a proper threshold level for warning? The threshold value needs to be selected to avoid frequent nuisance alarm while if still leaves enough time for the driver to react. The threshold value depends on the specific systems, including the vehicle, driver, and roadway characteristics. For example, the meandering behavior of drivers is an important factor to be considered, in order to reduce nuisance alarms. From several experimental studies [65], the lane position variance of human drivers is found to be between 0.24 and 0.61 ft^2. Intelligent algorithms to reduce the negative effect of spontaneous lane position offset to nuisance alarms are thus needed. Results reported in [66] show that a TLC of 1 s was able to prevent 92% of passenger car crashes and 60% of heavy vehicle crashes with less than 1 nuisance alarm per hour. A higher threshold value will improve crash prevention rate at the price of increased false alarm and nuisance alarm rates. Threshold value much lower than 1.0 s will leave little time for the driver to react and thus may not be advisable.

When it comes to the selection of warning mechanisms, no definite conclusion has emerged. In [67], through extensive simulator tests at the Daimler–Benz driving simulator and on-the-road experiments, it was found that the most desirable warning mechanism is through the use of simulated rumble strip sound. In [50], three warning concepts were tested and compared: a haptic warning with a slightly oscillating steering wheel; a corrective, haptic warning with an additional torque leading back to the middle of the lane; and an acoustic warning, the rumble strip sound. Eighteen professional truck drivers were interviewed after their 30 min test drive, and the acoustic warning was found to be the preferred mechanism. In [68], 28 drivers were asked to test drive a vehicle with both haptic feedback and acoustic feedback mechanisms. The acoustic feedback was found to be annoying. Sato et al. [69] found that pulse-like steering torque serves as an effective warning method and reduces the lateral deviation of the vehicle significantly. Suzuki and Jansson [70] argued that warning effectiveness and ease of recognition (of the warning) should be discussed simultaneously. By using a driving simulator, they found steering vibration to be the best mechanism, measured by steering reaction time and maximum steering angle. These findings, of course, depend on the specific design and refinement of the individual algorithms and selection of critical threshold values (timing and magnitude of warning). Much research is needed until a practical system can be assembled.

39.5 CONCLUSIONS

This chapter provides a review on recent developments and the current state-of-the-art approaches of vehicle lateral control systems. Some very important works in this immensely large and active field may have been left out. The realization of many of the ideas described in this chapter depends, to a great extent, on the progress of underlying enabling technologies, such as sensing systems, new concept of system architecture, and consideration of dominant system uncertainties. In addition, the associated economic (cost/benefit) and legal issues will continue to be important factors to be considered, especially for the U.S. market. Finally, the increasingly pressing issue of identifying sustainable future ground transportation methods cannot be ignored. If and when a big change of personal transportation mode occurs, the lateral/handling control problems may also become quite different. In any case, the lateral control of ground vehicles is an exciting and active area. The author feels quite honored to have the opportunity to know and work with many of these excellent researchers.

ACKNOWLEDGMENTS

This work contains excerpts or ideas from a couple of papers coauthored by a few former students, colleagues, and collaborators: Eric Tseng, Yigit Ungoren, Dave Leblanc, Chia-Shang Liu, and Nikhil Mudaliar.

REFERENCES

1. Chee, W., Tomizuka, M., Patwardhan, S., and Zhang, W.-B. Experimental study of lane change maneuver for AHS applications, in *Proceedings of the American Control Conference*, Seattle, WA, pp.139–143, 1995.
2. Tan, H.S., Guldner, J., Chen, C., Patwardhan, S., and Bougler, B. Lane changing with look-down reference systems on automated highways, *IFAC Journal, Control Engineering Practice*, 8(9), 1033–1043, September 2000.
3. Hatipoglu, C., Özgüner, U., and Redmill, K.A. Automated lane change controller design, *IEEE Transactions on Intelligent Transportation Systems*, 4(1), 13–22, March 2003.
4. Gardels, K. *Automatic Car Controls for Electronic Highways*, GMR, General Motors Corporation, Warren, MI, Report GMR-276, June 1960.
5. Cardew, K.H.F. *The Automatic Steering of Vehicles—An Experimental System Fitted to a DS 19 Citroen Car*, Road Research Lab, Crowthorne, Berkshire, U.K., 1970.
6. Fenton, R.E., Melocik, G.C., Olson, K.W. On the steering of automated vehicles: Theory and experiment, *IEEE Transactions on Automatic Control*, AC-21(3), 306–315, June 1976.
7. Fenton, R.E., Selim, I. On the optimal design of an automotive lateral controller, *IEEE Transactions on Vehicular Technology,* 37(2), 108–113, May 1988.
8. Peng, H. and Tomizuka, M. Preview control for vehicle lateral guidance in highway automation, *ASME Journal of Dynamic Systems, Measurement and Control*, 115(4), 679–686, December 1993.
9. Peng, H., Zhang, W., Tomizuka, M., and Shladover, S. A reusability study of vehicle lateral control system, *Vehicle System Dynamics*, 23(4), 259–278, May 1994.
10. Guldner, J., Tan, H., and Patwardhan, S. Analysis of automatic steering control for highway vehicles with look-down lateral reference systems, *Vehicle System Dynamics*, 26, 243–269, 1996.
11. Chen, C. and Tomizuka, M. Lateral control of commercial heavy vehicles, *Vehicle System Dynamics*, 33, 391–420, 2000.
12. http://www.ahsra.or.jp/
13. Hiramatsu, K. et al. *Design Principles of the Advanced Safety Vehicle*, 7th edn., World Congress on ITS, 2000.
14. Broggi, A., Bertozzi, M., Fascioli, A., and Conte, G. *Automatic Vehicle Guidance—The Experience of the ARGO Autonomous Vehicle*, World Scientific, Singapore, ISBN 981–02–3720–0.
15. http://www.phileas.nl/
16. Thorpe, C., Hebert, M., Kanade, T., and Shafer, S. Vision and navigation for the Carnegie-Mellon Navlab, *IEEE Transactions on PAMI*, 10(3), 401–412, 1988.

17. Dickmanns, E.D. An advanced vision system for ground vehicles, *The 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems*, Austria, April 2003.
18. Jaehne, B., Haubenecker, H., and Geibler, P. (eds.) *Handbook of Computer Vision and Applications*, Vol. 3, Academic Press, San Diego, CA, 1999.
19. Kato, S. and Tsugawa, S. Cooperative driving of autonomous vehicles based on localization, inter-vehicle communications and vision systems, *JSAE Review* 22, 503–509, 2001
20. Peng, H., Zhang, W., Tomizuka, M., and Arai, A. Magnetic-marker based lane keeping: A robustness experimental study, in *Automotive Display Systems and IVHS*, Society of Automotive Engineers, Warrendale, PA, 1993. Also in the *SAE Transactions—Journal of Passenger Cars*, 102–106, 1993.
21. Tan, H.-S., Bougler, B., and Zhang, W.-B. Automatic steering based on roadway markers—From highway driving to precision docking, *Vehicle System Dynamics*, 37(5), 315–339, March, 2002.
22. Farrell, J., Tan, H.-S., and Yang, Y. Carrier phase GPS-aided INS based vehicle lateral control, *ASME Journal of Dynamics Systems, Measurement, and Control*, 125(3), 339–353, September 2003.
23. Alexander, L. and Donath, M. Differential GPS based control of a heavy vehicle, in *Proceedings of the IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems*, Tokyo, Japan, pp. 662–667, October 1999.
24. http://www.ahsra.or.jp/eng/c04e/7th/08pos_e/11pos3_e.htm\
25. Kawazoe, H. et al. Development of a lane-keeping support system, SAE World Congress and Exhibition, SAE#2001–01–0797.
26. Mineta, K., Unoura, K., and Ikeda, T. Development of a lane mark recognition system for a lane keeping assist system, SAE World Congress and Exhibition, SAE# 2003–01–0281.
27. Redmill, K.A. et al. A lane tracking system for intelligent vehicle applications, *IEEE Intelligent Transportation Systems Conference*, Oakland, CA, pp.273–279, August 2001.
28. Chapuis, R., Aufrere, R., and Chausse, F. Accurate road following and reconstruction by computer vision, *IEEE Transactions on Intelligent Transportation Systems*, 3(4), 261–270, December 2002.
29. MacAdam, C.C. An optimal preview control for linear systems, *Journal of Dynamic Systems, Measurement and Control*, 102, 188–190, September, 1980.
30. MacAdam, C.C. Application of an optimal preview control for simulation of closed-loop automobile driving, in *Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics*, SMC-11, pp. 393–399, June 1981.
31. Shimakage, M. et al. *Design of Lane-Keeping Control with Steering Torque Input for a Lane-Keeping Support System*, SAE World Congress and Exhibition, SAE#2001–01–0480.
32. Iihoshi, A., *Driver Assistance System (Lane Keep Assist System)*, WP-29 ITS Round Table, Geneva, Switzerland, February 2004.
33. Mouri, H., Satoh, S., Furusho, H., and Nagai, M., Investigation of automatic path tracking using an extended Kalman filter, *JSAE Review* 23, 61–67, 2002.
34. Senger, K.H and Kortum, W. Investigations on state observers for the lateral dynamics of four wheel steered vehicles, *The Dynamics of Vehicles on Roads and Tracks, Supplement to Vehicle System Dynamics*, 18, 515–527, 1989.
35. Kaminaga, M. and Naito, G. Vehicle body slip angle estimation using an adaptive observer, in *Proceedings of the 4th International Symposium on Advanced Vehicle Control (AVEC)*, Nagoya, Japan, pp. 207–212, September 14–18, 1998.
36. Liu, C. and Peng, H. A state and parameter identification scheme for linearly parameterized systems, *ASME Journal of Dynamic Systems, Measurement and Control*, 120(4), 524–528, December 1998.
37. Farrelly, J. and Wellstead, P. Estimation of vehicle lateral velocity, in *Proceeding of the 1996 IEEE International Conference on Control Applications*, Dearborn, MI, pp.552–557, September 15–18, 1996.
38. Fukada, Y. Estimation of vehicle side-slip with combination method of model observer and direct integration, in *Proceedings of the 4th International Symposium on Advanced Vehicle Control (AVEC)*, pp. 201–206, Nagoya, Japan, September 14–18, 1998.
39. Nishio A. et al. *Development of Vehicle Stability Control System Based on Vehicle Sideslip Angle Estimation*, 2001 SAE World Congress, SAE paper 2001–01–0137.
40. Hac, A. and Simpson, M.D. Estimation of vehicle side slip angle and yaw rate, 2000 SAE World Congress, SAE paper 2000–01–0696.
41. Tseng, H.E. Dynamic estimation of road bank angle, in *Proceedings of the 5th International Symposium on Advanced Vehicle Control (AVEC)*, Ann Arbor, MI, pp. 421–428, August 22–24, 2000.
42. Mudaliar, N., LeBlanc, D., and Peng, H. Linear estimator for road departure warning systems, in *Proceeding of the 2004 American Control Conference*, Boston, MA, June 2004.

43. Ungoren, A.Y., Peng, H., and Tseng, H. A study on lateral speed estimation methods, *International Journal of Vehicle Autonomous Systems*, 2(1/2), 126–144, 2004.
44. Feng, K.-T., Tan, H.-S., and Tomizuka, M. Automatic steering control and validation of vehicle lateral motion with the effect of roll dynamics, in *Proceedings of American Control Conference*, Philadelphia, PA, pp. 2248–2252, June 1998.
45. Tan, H.S. Develop precision docking function for bus operation, *California PATH Research Report*, UCB-ITS-PRR-2003–11, March 2003.
46. Hiramatsu, S. et al. Rearview Camera Based Parking Assist System with Voice Guidance, 2002 SAE World Congress, SAE paper 2002–01–0759.
47. Katsuno, T. et al. Development of a parking assist system with voice guidance, *2002 JSAE Annual Congress*, JSAE Technical Paper No. 20025067 (also known as SAE paper 2002–08–0047).
48. Najm G et al. Analysis of light vehicle crashes and pre-crash scenarios based on the 2000 general estimates system, Department of Transportation Report HS 809 573, Washington, DC.
49. LeBlanc et al. CAPC: An implementation of a road-departure warning system, in *Proceedings of the 5th IEEE International Conference on Control Applications*, Dearborn, MI, September 1996.
50. Bishel, R. et al. Lane departure warning for CVO in the USA, SAE Publications SP-1401 Electronics for Trucks and Buses, SAE paper#982779.
51. Lin, C.-F., Ulsoy, A.G., and Leblanc, D.J. Vehicle dynamics and external disturbance estimation for vehicle path prediction, *IEEE Transactions on Control System Technology*, 8(3), 508–518, May 2000.
52. Hahn, J.O., Rajamani, R., and Alexander, L. GPS-based real-time identification of tire–road friction coefficient, *IEEE Transactions on Control System Technology*, 10(3), 331, May 2002.
53. Eichhorn, U. and Roth, J. Prediction and monitoring of tyre/road friction, in *Proceedings of FISITA*, London, U.K., June 1992.
54. Ito, M., Yoshioka, K., and Saji, T. Estimation of road surface conditions using wheel speed behavior, SAE paper no. 9438826.
55. Pal, C., Hagiwara, I., Morishita, S., and Inoue, H. Application of neural networks in real time identification of dynamic structural response and prediction of road-friction coefficient from steady state automobile response, SAE paper no. 9438817.
56. Pasterkamp, W.R. *The Tyre as Sensor to Estimate Friction*, Delft University Press, Delft, the Netherlands, 1997.
57. Gustafsson, F. Monitoring tire–road friction using the wheel slip, *IEEE Control Systems Magazine*, 18, 42–49, August 1998.
58. Liu, C. and Peng, H. Road friction coefficient estimation for vehicle path prediction, *Vehicle System Dynamics*, 25(Suppl.), 413–425, 1996.
59. Umeno, T. et al. Estimation of tire-road friction using wheel speed vibration, in *Proceedings of the 2002 AVEC Conference*, Hiroshima, Japan.
60. Umeno, T. et al. Estimation of tire-road friction by tire rotational vibration model, *R&D Review of Toyota CRDL*, 37(3), 53–38, http://www.tytlabs.co.jp/office/elibrary/ereview/rev373epdf/e373_053umeno.pdf
61. Umeno et al. Observer based estimation of parameter variations and its application to tyre pressure diagnosis, *Control Engineering Practice* 9, 639–645, 2001.
62. Godthelp, H., Milgram, P., and Blaauw, G.J. The development of a time-related measure to describe driver strategy, *Human Factors*, 26(3), 257–268, 1984.
63. Batavia, P.H. *Driver-Adaptive Lane Departure Warning Systems*, Ph.D. Dissertation, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1999.
64. Pilluti, T. and Ulsoy, A.G. Fuzzy-logic-based virtual rumble strip for road departure warning systems, *IEEE Transactions on Intelligent Transportation Systems*, 4(1), 1–12, March 2003.
65. Pape, D.B. et al. Performance considerations for run-off-road countermeasure systems for cars and trucks, SAE paper#1999–01–0820.
66. Pomerleau, D. SafeTRAC: Research to deployment, http://www.ivsource.net/public/ppt/RORAssistWare_Natl_IVI_Mtg2000.ppt
67. Ziegler, Walter et al. Computer vision on the road: A lane departure and drowsy driver warning system, SAE Paper#952256 E6.
68. Pohl, J. and Ekmark, J. Development of a haptic intervention system for unintended lane departure, SAE paper# 2003–01–0282.
69. Sato, et al. A study on a lane departure warning system using a steering torque as a warning signal, *Proceedings of the 1998 AVEC Conference, Paper #9837111*, Nagoya, Japan, 1998.
70. Suzuki, K. and Jansson, H. An analysis of driver's steering behaviour during auditory or haptic warnings for the designing of lane departure warning system, *JSAE Review*, 24, 65–70, 2003.

40 Longitudinal Control

Paul Fancher and Charles MacAdam

CONTENTS

40.1 Introduction 1549
40.2 Basic Structure of Longitudinal Control Systems 1550
40.3 Insight into System Design and Evaluation 1551
40.4 Adaptive Cruise Control 1555
40.4.1 Adaptive Cruise Control Using the Elements of a CCC System 1555
40.4.2 ACC Using Acceleration Control 1560
40.4.3 Unification Approach Applied to ACC with a Velocity Actuator 1562
40.4.4 ACC Systems with an Acceleration Actuator and No Available Acceleration Sensor 1562
40.4.5 Summary of Related ACC Designs 1563
40.4.6 Evaluation of ACC Systems 1565
40.5 Emulation of the Longitudinal Control Properties of Human Driving 1567
40.5.1 Nonlinear Time Gap ACC Systems 1569
40.6 Forward Collision Warning 1570
40.7 Automated Highway Systems 1572
40.8 Closure to the Overview of Basic Concepts 1575
40.9 Advanced Topics 1576
40.9.1 Driver Closed-Loop Braking Performance under Limit Conditions 1576
40.9.2 Driver Headway Control Behavior 1576
40.9.3 Sliding Control 1576
40.9.4 Driver Longitudinal Control Behavior When Entering Curves 1576
40.9.5 Nonlinear Control 1576
40.9.6 Control Hysteresis and Saturation 1576
40.9.7 Control Delays 1577
40.9.8 Platoons and String Stability 1577
40.9.9 Artificial Intelligence Methods Applied to Longitudinal Control 1577
References 1577

40.1 INTRODUCTION

This chapter provides basic insights into the longitudinal control (LC) of roadway vehicles. It relates fundamental physical concepts involving desired range clearance, speed control, and acceleration/deceleration performance to control system features. The systems discussed extend classical notions of vehicle dynamics to the dynamics of vehicles with systems for increasing driver comfort and convenience on the one hand, and/or for improving efficiency, stability, or active safety on the other. These systems provide driver assistance and extend the domain of vehicle control applications by employing remote sensing and information gathering concerning the driving environment outside of the motor vehicle itself. Nevertheless, LC has been a driver activity, and current driver assistance systems involve driver decision-making and supervision at the very least. To the extent that we know how people control vehicles, we will use that knowledge in evaluating the performance of driver assistance systems.

This chapter is aimed at complementing other sections in this handbook by providing basic insights into LC systems in a manner that will aid either in the design and creation of new systems or in the evaluation of currently existing systems. In keeping with these aims, simplified analyses and models are used for coping with the complexity of real systems and the features of their hardware and software. After certain fundamental concepts are presented and discussed, the chapter concludes with annotated literature sources that are pertinent to advanced treatment of LC topics.

40.2 BASIC STRUCTURE OF LONGITUDINAL CONTROL SYSTEMS

Figure 40.1 is a generic block diagram containing the basic functional elements of an LC system. These elements are the traffic–roadway subsystem, the command and control unit (which may be driver or vehicle based), and the vehicle and its actuators. These three subsystems constitute the elements of an LC system that employs remote sensing to perceive the driving situation. This same form of diagram has been applied to manual driving in which the driver performs the sensing, perceiving, deciding, and controlling (commanding) functions shown in Figure 40.1. In addition, this form of diagram could be applied to lateral control systems as well, but the variables, as indicated here, pertain to the inputs and outputs of LC elements.

Specifically, the pertinent variables are as follows:

V, vehicle velocity, is the speed of a controlled (LC-equipped) subject vehicle.
V_p is the velocity time history of the motion of the vehicle in front of the subject vehicle.
R, range clearance, is the inter vehicular distance as shown in Figure 40.2. The range clearance is the difference between the longitudinal position X_p of the preceding vehicle and the position X of the subject vehicle with respect to a reference location ($R = X_p - X$).
R_{dot}, range rate, is dR/dt, the time rate of change of R and the relative velocity between a vehicle pair. See Figure 40.2. $R_{dot} = V_p - V$, and R equals the integral of R_{dot}.
U, control command, is the control input used to command vehicle speed and/or acceleration.
V_{dot} is dV/dt, the acceleration/deceleration of the subject vehicle.
V equals the integral of V_{dot}.

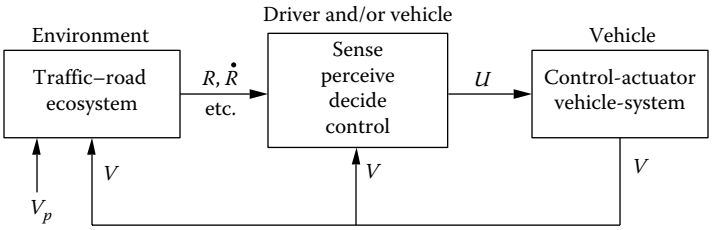


FIGURE 40.1 Basic diagram of a LC system.

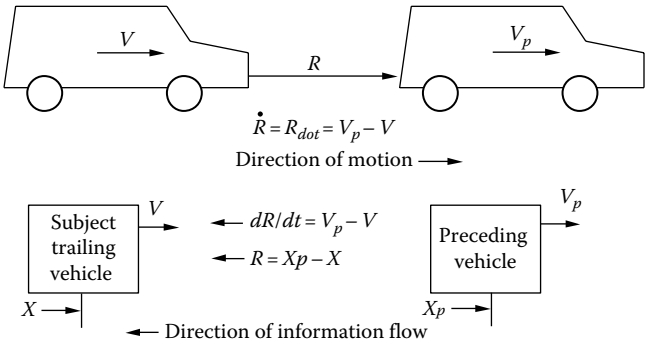


FIGURE 40.2 Illustrated variables.

Figure 40.2 illustrates the geometric and kinematic relationships between a subject (LC) vehicle and its preceding vehicle. For many LC applications, the simple kinematic relationship, describing the relative velocity between the subject vehicle and its preceding vehicle, is a crucial factor in understanding the driving situation, viz.,

$$R_{dot} = V_p - V \tag{40.1}$$

If V_p is greater than V, R_{dot} is positive and range is increasing. In this case, the vehicles are separating. If V is greater than V_p, R_{dot} is negative and range is decreasing. In this situation, the trailing LC vehicle is closing on the preceding vehicle. To maintain a desired range, R_{dot} must be zero. If R_{dot} remains close to zero and averages approximately zero over an extended period of time, the situation is referred to as following. Separating, closing, and following are the basic roadway situations considered in the longitudinal aspects of driving.

The hardware employed in LC systems consists of sensors, command and control units, actuators, and, of course, the vehicle itself. Generally, sensors are required for measuring range R, range rate R_{dot}, and velocity V. Also the acceleration V_{dot} (also symbolized as A) is often measured since it is used in certain control laws to formulate acceleration commands. The command and control units use the outputs of the sensor units to compute control commands (generally represented by U in Figure 40.1). The actuation mechanisms related to LC control are in the braking system and engine/driveline system of the vehicle. These mechanisms are viewed as providing the tire forces that decelerate or accelerate the vehicle. We will use simplified models to represent the vehicle and its braking and accelerating capabilities. Other sections of this handbook treat those topics in detail.

40.3 INSIGHT INTO SYSTEM DESIGN AND EVALUATION

Ideas involved with system design and system evaluation are used in this chapter to help organize the material and to establish an overall context supporting the discussion. Figure 40.3 presents an overview of the design and evaluation process [1]. The levels of abstraction in going from functional purpose and concept-ideas to a real system are indicated in Figure 40.3. Depending on the choices made at any of the levels of abstraction shown in Figure 40.3, a wide variety of final system configurations and capabilities is possible. A key area in developing a unified design, which fits together well, is the creative process of unifying function with form and form with function. We will indicate how unity may be achieved by discussing in a simplified way the characteristics of the components used in developing LC systems.

Clearly, commercial systems are often designed with particular vehicles in mind, and the sensors may have been previously chosen. In this sense, component characteristics represent constraints on

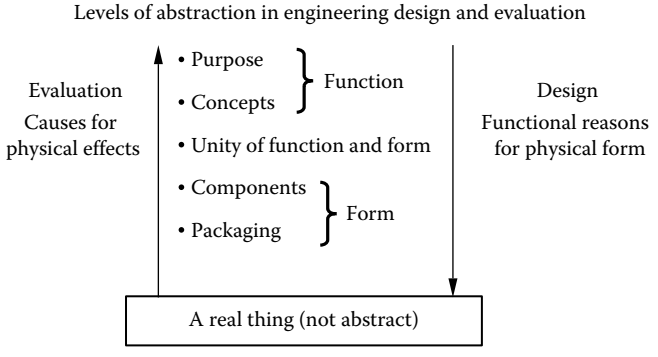


FIGURE 40.3 From abstract functional concepts to the creation of a real system.

the design options available in creating a system. In practice, the order of the levels of abstraction shown in Figure 40.3 does not imply a time sequence of events in developing or evaluating a system. For example, the existence of new types of sensors for measuring range and range rate may be the first step that initiated the design process for a particular adaptive cruise control (ACC) system. Generally, designers may jump around from one level of abstraction to another in the development of an LC system as they iterate toward a design.

Although conventional cruise control (CCC) is considered to be a system that was developed before the new classes of driver assistance systems were introduced, we wish to consider it here to illustrate how to employ the higher levels of abstraction in discussing LC systems. In addition, there are a number of basic ideas that are well illustrated by this example. After all, CCC, engine control systems, and antilock braking systems were precursors to the electronic control systems used in intelligent vehicle systems today. (These specific systems are treated in detail in Chapter 39.)

Usually, our discussions will be top down, emphasizing the higher levels of abstraction in our examples. The functional purpose of the CCC system is to maintain a driver-selected traveling velocity. This is called the set speed "V_{set}" for the CCC system. The driver is to be responsible for turning the system on and off. If a slower moving vehicle or some other obstacle obstructs the driver's path, the driver is to brake, and this will turn the CCC function off, as well as control the situation.

With regard to functional concepts, information flow diagrams and system block diagrams provide convenient means for expressing basic ideas. Figure 40.4 illustrates the basic ideas underlying a CCC system.

The input is V_{set}, and it is compared to the actual speed V to form an error signal for the control system. This indicates the need for a sensor to measure V. At this level of abstraction, we treat the sensor as being accurate and ignore any sensor dynamics or errors involved. Those items would be something to look at in component engineering. If we choose a simple model for the vehicle and its braking and accelerating actuators, the control system concepts would appear as shown in Figure 40.4. The equations pertaining to Figure 40.4 are as follows:

For the vehicle and actuators,

$$M_e V_{dot} = F_c + F_R + F_G \tag{40.2}$$

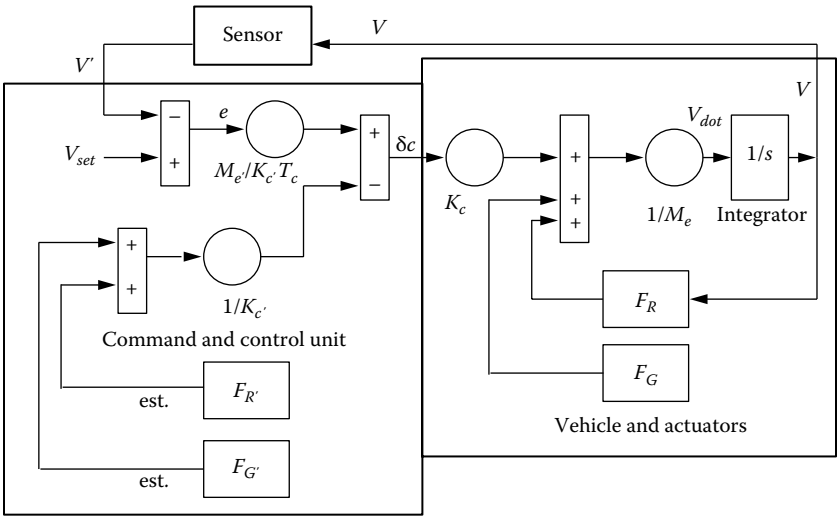


FIGURE 40.4 Concept diagram for a CCC system.

where

M_e is the effective mass of the vehicle and drive-train and engine or brake components and tires
V_{dot} is the vehicle acceleration/deceleration
F_c is the actuation force (which will be discussed further)
F_R is the natural retardation due to rolling resistance and aerodynamic drag
F_G is the influence of the highway grade on longitudinal acceleration

In Figure 40.4, the actuation force is represented by the following simplified equation:

$$F_c = K_c \delta_c \tag{40.3}$$

where

K_c is a gain factor
δ_c is the control input signal

Equation 40.3 is an abstraction used at the conceptual level. It is well removed from the real thing.

The simplifications used in these abstract equations indicate that detailed considerations will need to be treated before a real operating system can be established. For example, the actuation force F_c is really a combination of the longitudinal tire force produced by the braking system and that produced by the engine and drive-train. These systems have their own gains, nonlinearities, and dynamics. There are other interesting abstractions associated with the effective mass and its relationship to velocity, the natural retardation, and the grade influences that will need to be specifically treated before a system meeting the functional purpose of the CCC system is realized.

The control and command unit in Figure 40.4 is represented by the following equations:

$$e = V_{set} - V \tag{40.4}$$

Once the driver selects V_{set}, it is considered to be constant and, hence, $de/dt = -V_{dot}$. And the following equation can be used in planning a speed regulator:

$$T_c \frac{de}{dt} + e = -T_c V_{dot} + (V_{set} - V) = 0 \tag{40.5}$$

Equations 40.4 and 40.5 are of the form of a simple proportional control system. However, if we have knowledge of how the vehicle and its actuators perform, we can unify the design with regard to function and form. We can use knowledge of the vehicle and actuator components to determine the actuator input δ_c, viz., by combining Equations 40.5, 40.2, and 40.3 algebraically, one obtains

$$\delta_c = \left(\frac{M_e'}{K_c' T_c} \right) e - \frac{\left(F_R' + F_G' \right)}{K_c'} \tag{40.6}$$

The primed quantities in Equation 40.6 represent estimations of the parameters and the natural retardation and grade factors pertaining to the actual vehicle situation. At least in concept, Equation 40.6 represents the idea of compensating for the way the vehicle works in order to get the vehicle to do what we want it to do. The idea is to arrange the CCC system such that the following differential equation provides a reasonable approximation of the system's performance:

$$T_c V_{dot} + V = V_{set} \tag{40.7}$$

Examination of Figure 40.4 indicates that we plan to cancel the effects of M_e, K_c, F_R', and F_G'. This plan may not be straightforward to accomplish. In practice, the effective mass M_e may depend

upon V in a complicated manner. The natural retardation will depend upon vehicle weight as loaded and ambient wind conditions as well as vehicle velocity. In component design, there is an issue of whether measuring wind speed and current weight is worth the cost and complexity. And highway grade changes continuously, depending upon the local terrain. Whether or how the system will adjust for grade is a design choice.

There are a number of practical design choices to be resolved in the design of the command and control unit. For example, simple proportional control might be augmented with some integral control to correct for the effects of long sustained grades or constant head or tail winds. The control system might involve parameter identification features that find good values for approximating M_e, K_c, F_R, and F_G or the parameters of a more complete model of the vehicle and its actuators. An adaptive control approach might be used. Sliding mode control might be applied with throttle valves and brake valves being used to approximate the gain and quickness needed for use with sign(e) types of terms. It is beyond the resources of this article to go further into these matters. See control system texts such as [2] for explanations of the theory of possible approaches. For references in the literature dealing with CCC systems, see [3] for examples.

The sensor component may introduce system deficiencies through magnitude errors, dynamic effects, and delays. In the CCC case treated here, forward velocity is the main variable required to be measured in the basic concept of the system. Since measurement and display of velocity have been traditional in highway vehicles, there may well be velocity information available for use in the CCC application. Modern powertrain controllers, antilock braking systems, and other features of modern vehicles may have sophisticated electronic means for assessing forward speed.

(A modern powertrain controller may have provision for engine and transmission control systems. Transmission controls use measurements of throttle angle and vehicle speed in determining the shift point and lockup operation point. Engine control systems provide a number of functions and use sensor signals for measuring a variety of quantities including vehicle speed, longitudinal slip, and vehicle acceleration. Antilock braking systems use measurements of wheel speed to modulate brake pressures in a manner that aids in preventing wheel lockup while maintaining good stopping performance. For further description of these systems and references, see [3] or other college texts.)

These types of systems operate using digital communication. For example, velocity data might be updated every 0.1 s with a resolution of approximately 0.2 m/s (0.5 mph). This means that for a one resolution-unit change in the velocity value, the output of a zero-order hold would change by 0.2 m/s. When the sampling period is taken into account, this translates approximately into 0.2 m/s/0.1s of acceleration, that is, 0.2 g. For CCC applications, this might be acceptable if T_c is large enough, but for ACC and tighter speed control situations, such jumps in acceleration commands may cause driver dismay and discomfort. A reduction in the level of these pseudo-acceleration pulses to 0.04 g would be much better. Although it might cause some design difficulties, a resolution of 0.04 m/s (i.e., 0.1 mph) in velocity could improve the situation.

Also due to the introduction of electronic communications in vehicles and the associated vehicle buses, information flow diagrams are often used in illustrating the connections between components. Besides giving an overview of the components included in a particular system, these diagrams provide a basis for seeing synergistic connections between other existing, (or recently proposed systems) and the particular system currently under design consideration. Figure 40.5 is an example showing an information flow diagram for the CCC design previously portrayed in Figure 40.4.

The communication bus (also known as "vehicle bus") clearly plays a central role in Figure 40.5. (For an explanation of the meaning of this term, look up "vehicle bus" at wikipedia.org on the World Wide web. In particular, see the Background section.) To illustrate possible synergy with other systems, a global positioning system (GPS) has been added. It is shown at the lower right of the diagram. Otherwise, the components and their interconnections follow from the previous discussions. As an example for this case, information from the GPS is to be applied in estimating the grade. The vertical velocity determined from GPS measurements is combined with the measured forward velocity to estimate grade. Perhaps it would be just as easy to use the GPS measurement of horizontal velocity, but that is not the plan chosen here.

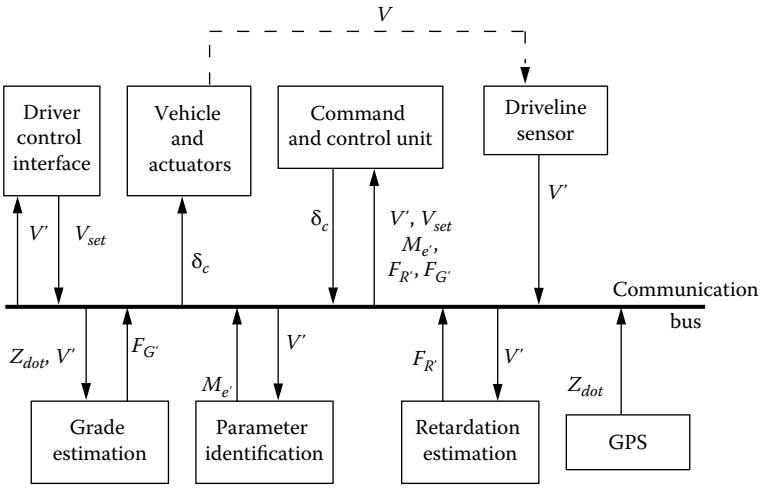


FIGURE 40.5 Information flow diagram for a CCC system.

This may appear to be a far-fetched way to estimate grade. A far simpler solution is just to ignore grade and let the driver work around it. However, if the vehicle already has a GPS that puts the needed information on the communication bus, the added cost for estimating grade may not be very large.

From a mechanical dynamics standpoint, the introduction of electronic communications (vehicle buses) opens up many opportunities for employing sensors, actuators, and control units without the need for linkages, mechanical cables, and hydraulic and pneumatic lines. And there is greater flexibility in the types of control features that digital controllers can readily provide.

40.4 ADAPTIVE CRUISE CONTROL

40.4.1 Adaptive Cruise Control Using the Elements of a CCC System

Recent ISO [4] and SAE [5] standards for ACC systems describe ACC as an extension of CCC. The functional purpose of the ACC system is to maintain a desired range clearance between an ACC-equipped vehicle and its preceding vehicle, when one is present. If there are no preceding vehicles within a reasonable range, the system is to perform as a CCC system.

In the example presented next, the extension from CCC to ACC is envisioned as employing the features of an existing CCC system in the conceptual design of the ACC system. For pragmatic and practical purposes, the CCC system is considered to be one of the components of the ACC system. See Figure 40.6 showing a CCC unit in the lower right area of the diagram.

Figure 40.6 resembles Figure 40.5 except that there are separate boxes for ACC and CCC command and control units and there are sensors for measuring pertinent features of the driving scene (labeled R' and R'_{dot}). The scene sensors are presumed to work well so that $R' = R$ approximately and $R'_{dot} = R_{dot}$ approximately and $dR'/dt = R'_{dot}$ approximately. The command input to the CCC unit comes from the ACC unit, and it is called V_c, where this command may ask for a speed that is less than the value of V_{set} chosen by the driver. In this system, V_{set} is the maximum speed that the ACC control will command. The ACC controller may command less than V_{set} in order to satisfy the range clearance purpose of the ACC system.

A number of the boxes involved in the CCC system diagram (Figure 40.5) are omitted in Figure 40.6 because it is assumed that the CCC system is well approximated by Equation 40.7 with V_{set} replaced by V_c, which is the output of the ACC controller, viz.,

$$T_c V_{dot} + V = V_c \tag{40.8}$$

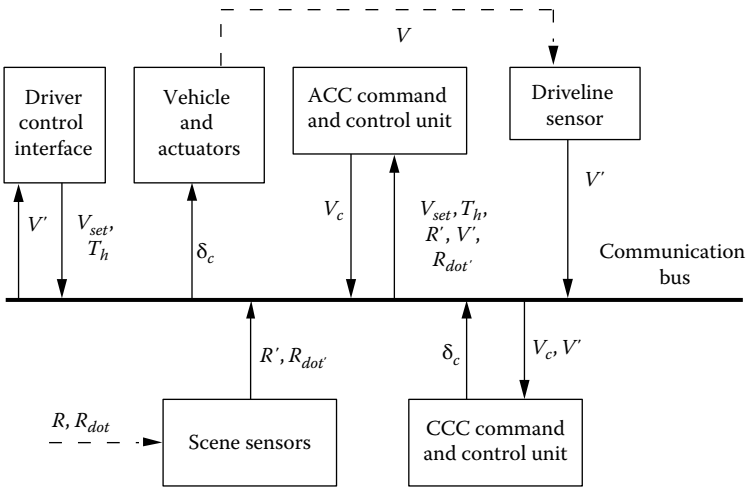


FIGURE 40.6 Information flow diagram for an ACC using CCC directly.

A point, which is subtly presented by Figure 40.6, is that there are inner (sub) and outer (main) control loops corresponding to the functions of the CCC controller and the ACC controller, respectively.

The ACC system diagram in Figure 40.7 makes these loops much more apparent, and it provides insight into the concepts underlying the design of a particular ACC controller. The inner sub loop is identified by the velocity feedback arrangement in the CCC unit. The outer main loop is identified by the velocity feedback path that goes through the traffic kinematics and then through the remote sensing block into the ACC command and control unit. (Information and concepts concerning traffic kinematics are indicated in the traffic kinematics block in Figure 40.7.)

The equations describing the nature of the ACC system portrayed in Figure 40.7 are as follows:

Equation 40.8, describes the function of the CCC unit (i.e., $T_c\, V_{dot} + V = V_c$).

As in Equation 40.1, traffic kinematics matters are described by $R_{dot} = V_p - V$.

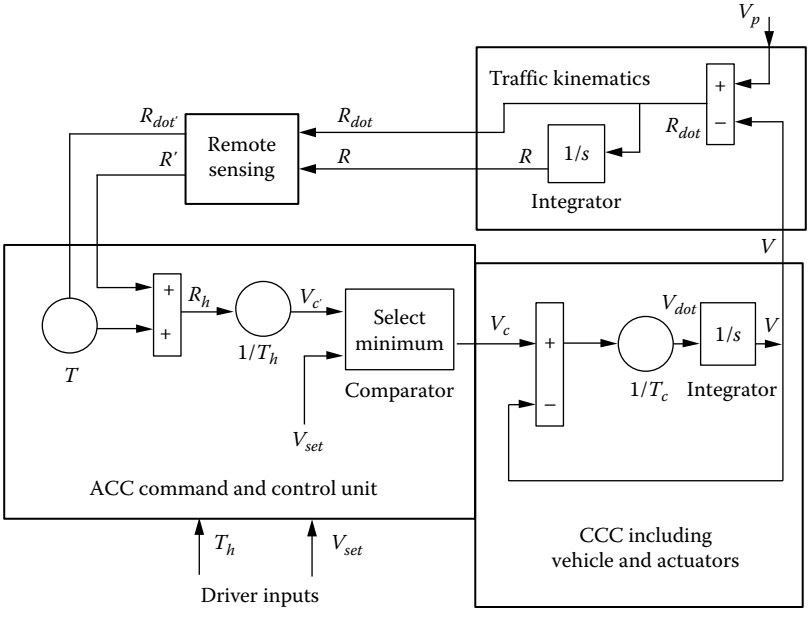


FIGURE 40.7 Concept diagram for an ACC system using CCC.

The functions of the ACC command and control unit may be expressed as follows:

$$R_h = R + T\ R_{dot} \tag{40.9}$$

R_h is the desired range for steady following of a preceding vehicle. And

$$V_c' = \frac{\left(R + T\ R_{dot}\right)}{T_h} \tag{40.10}$$

T_h is the desired time gap chosen by the driver.

The driver also chooses a value for V_{set}. V_{set} is the maximum speed that the driver chooses to use on the road currently being traveled. A comparator–limiter is employed so that V_c is always less than or equal to V_{set}.

There are concerns over the values of V_{set} and T_h that are appropriate for use in ACC systems. In particular, the maximum value of V_{set} and the minimum value of T_h are currently considered issues. For example, a maximum value of V_{set} might have a design value of 85 mph (38 m/s) in the United States. Also the minimum value of V_{set} might be about 30 mph (13 m/s) since the ACC and CCC systems may be designed to quit functioning below its minimum speed. (Hence, the driver is required to drive manually in low-speed situations.) The time gap parameter may have a restricted range of settings from ~1.0 to 2.0 s, although many drivers follow at less than 1.0 s when driving manually at freeway speeds. The reasons given for these restrictions on V_{set} and T_h have to do with highway safety, but the issues are subtle and universally acceptable compromises are not easily justified.

The space of sensor measurements (R' and R'_{dot}) may be displayed in a range versus range rate phase plane diagram. This type of diagram is useful for explaining the reasoning behind Equations 40.9 and 40.10 used in the ACC controller. In Figure 40.8, the vertical axis is the range R and the horizontal axis is the range rate R_{dot}. Measured values of R and R_{dot} could fall anywhere in this R–R_{dot} space at any time, depending upon the relative position and relative velocity at that particular time.

The diagonal straight line shown in Figure 40.8 represents Equation 40.9. This line represents a linear first-order differential equation as displayed in this phase plane. The arrows on the line show the direction of motion along this line starting from any point on this line. The point at $R = R_h$ and $R_{dot} = 0$ is a stable equilibrium point since R_{dot} is negative for $R > R_h$ and R_{dot} is positive for $R < R_h$.

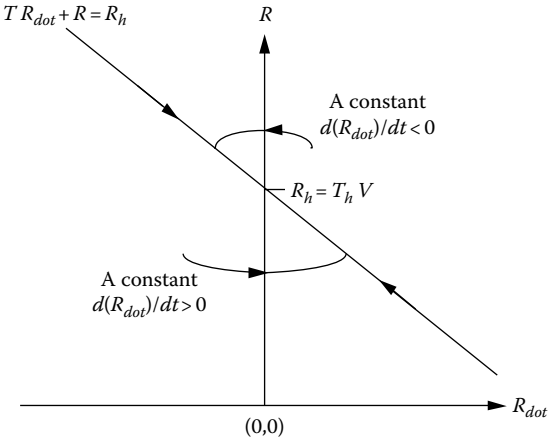


FIGURE 40.8 Phase plane for R and R_{dot}.

The slope of this line is equal to $-T$, where T is the time constant of the hypothetical first-order differential equation portrayed in Figure 40.8. If the current values of R and R_{dot} were to lie on this line, the ensuing trajectory would slide along the line $T\,R_{dot} + R = R_h$ (a sliding surface in the two-dimensional space of R and R_{dot}).

The point at $R = R_h$ and $R_{dot} = 0$ represents the desired range at a given speed, that is, we want $T_h\ V = R_h$. Equation 40.10 simply indicates that V_c' is the desired speed conforming to a constant time gap policy where T_h is the desired time gap. In other words, the value of R_h shown in Figure 40.8 changes with velocity, and consequently the line, $T\,R_{dot} + R = R_h$, changes its R- intercept as velocity changes. In this context, it may be said that the control philosophy is to ask for a speed V_c that will lead to what you want. Clearly, this requires knowing what you want to do. The line $T\,R_{dot} + R = R_h$ represents the plan for getting to R_h, which is where you want to be driving during steady following.

The two curved lines in Figure 40.8 represent lines of constant $d(R_{dot})/dt$. They are lines of constant relative acceleration. For cases in which $V_{p\,dot} = 0$ approximately, $d(R_{dot})/dt = -V_{dot}$ approximately and $d(R_{dot})/dt > 0$ represents braking. As indicated in Figure 40.8, braking (when the preceding vehicle is traveling at nearly constant speed) will cause range to decrease during closing and then cause separating to occur once V falls below V_p. For constant deceleration cases, the form of the curve is a parabola. The parabola for $d(R_{dot})/dt < 0$ is also shown in Figure 40.8, and it applies to the acceleration of the ACC vehicle when the preceding vehicle is traveling at approximately constant speed.

Implicit in this discussion is the idea that the driver takes control of the system when a close preceding vehicle is braking at anything approaching the control authority of the ACC system. Depending upon design characteristics, ACC systems may employ maximum deceleration authorities from 0.1 to 0.3 g approximately. Current commercially available systems tend to utilize the higher levels of control authority in setting system deceleration limits in order to extend the range of utility of the ACC system [6,7]. Acceleration/deceleration factors will be discussed later in connection with ACC systems based upon a desired acceleration/deceleration command structure.

However, braking commands figure into this ACC-using-CCC case. Once conditions reach the point (R and R_{dot} conditions) where throttle modulation is not sufficient to approach the desired dynamics implied by the $T\,R_{dot} + R = R_h$ line, light braking may be used to produce deceleration levels that will intersect the desired dynamics line from the bottom [8]. Constant acceleration lines can be used to intersect the desired dynamics line from the top of the line in the phase plane. In this sense, the desired dynamics line is both a place (in the phase plane) and a dynamics model. More information on phase plane dynamics is reported in Chapter 5. A basic idea here is to switch to the sliding mode rules when the R and R_{dot} conditions are close enough to the sliding surface (more information on sliding mode control is reported in Chapter 6). Although not immediately obvious, one of the aspects of this type of control is that the constant deceleration situations will mean that R falls below R_h in the process of approaching $R = R_h$, the desired following situation. These "undershoots" in range may cause concern and driver braking to avoid short ranges. And the propagation of increasing undershoots from vehicle to vehicle in a string will disrupt the integrity of the performance of the string such that the driver of the last vehicle in the string will need to intervene by braking to avoid crashing into the rear of the preceding vehicle. Nevertheless, systems of this type have been studied in field operational tests and shown to have favorable subjective ratings from lay drivers operating in naturalistic conditions during normal driving [9].

The dynamics of strings of vehicles running on highways is presented in Chapter 39. Ref. [10] treats matters of asymptotic string stability with mathematical rigor. That material addresses highway automation in which the goal is to automatically maintain short headway distances between vehicles regardless of vehicle speed. The following material is aimed at the safety concerns and traffic flow disruptions that may be caused by the dynamics of ACC systems.

From an analytical standpoint, it is interesting to examine the transfer function for the system portrayed in Figure 40.7. For $V_c < V_{set}$, the open-loop transfer function Y_o between R_{dot} and V is given by the following equation:

$$Y_o = \frac{V(s)}{R_{dot}(s)} = \frac{\left[T + \left(1/s\right)\right]}{\left[T_h\left(T_c s + 1\right)\right]} = \frac{\left(Ts + 1\right)}{\left(T_h s\right)\left(T_c s + 1\right)} \tag{40.11}$$

where s represents the Laplace transform variable used in the linear analysis of differential equations.

Examination of Equation 40.11 makes it apparent that $T = T_c$ is a special case in which

$$Y_o = \frac{1}{T_h s} \tag{40.12}$$

The zero/pole cancellation implied by making $T = T_c$ turns a more complicated open-loop transfer function into a simple gain and integrator. This case plays an important role in understanding what various choices of T mean when T_c is a design constraint determined by the CCC system chosen for use in the ACC system.

Based upon Equation 40.11, the closed-loop transfer function for the main loop is as follows:

$$\frac{V(s)}{V_p(s)} = Y_c = \frac{Y_o}{\left(1 + Y_o\right)} = \frac{\left(Ts + 1\right)}{\left(Ts + 1 + T_h s\left(T_c s + 1\right)\right)} \tag{40.13}$$

If

$$T = T_c, \quad Y_c = \frac{1}{\left(T_h s + 1\right)} \tag{40.14}$$

Equation 40.14 implies that if T provided enough lead to cancel the effects of the lag due to T_c, then the system would act like a simple first-order system. And there would be no overshoot or undershoot in response to abrupt changes in V_p (the speed of the preceding vehicle).

For the general case of Equation 40.13, the condition for assuring that the amplitude of disturbances will not increase going back along a string of like vehicles is $|V(s)/V_p(s)|^2 \leq 1.0$ for all s [11]. Upon replacing s by $j\omega$ and performing algebraic operations, the condition can be reduced to the following equivalent inequality for the system of Equation 40.13:

$$1 + T^2\omega^2 \leq (1 - T_c T_h \omega^2)^2 + (T + T_h)^2 \omega^2 \tag{40.15}$$

After simplification of Equation 40.15, it is found that the constraint on the parameters is most stringent at $\omega = 0$ and the result is

$$T > \frac{T_c - T_h}{2} \tag{40.16}$$

The point is that for smooth string behavior we need enough lead (T) to counteract the lag (T_c) associated with the CCC actuator. String behavior issues are critical in the discussions of traffic management and congestion alleviation.

40.4.2 ACC Using Acceleration Control

This case differs from the previous case in that the CCC velocity actuator is replaced by an actuator that accepts acceleration commands. The basic equation for this type of actuator is

$$\tau A_{dot} + A = A_c \tag{40.17}$$

The block diagram for the ACC system using (40.17) is presented in Figure 40.9.

As shown in Figure 40.9, the blocks involving the vehicle and actuator and the ACC command and control unit are different from those shown in Figure 40.7, which pertains to an ACC system using a velocity command structure. Clearly, the traffic kinematics block remains the same. The designer has no control over that. Figure 40.9 indicates explicitly that internal sensors for both the velocity V and the acceleration A, where $A = V_{dot}$, need to be used in the control concept envisioned for this system. The use of A in the design of this control algorithm is an important factor that allows us to unify the components of the system in an advantageous manner.

This unification is implemented through the equations for the acceleration command A_c' shown in the ACC command and control unit block in Figure 40.9. The conceptual reasoning behind that control rule (algorithm) is as follows:

Step 1: A special error function is created. The form of this function contains both the desired headway time gap and an acceleration term. The acceleration term is needed because the chosen actuator is acceleration based as indicated by Equation 40.17. The error function needs to contain a link to

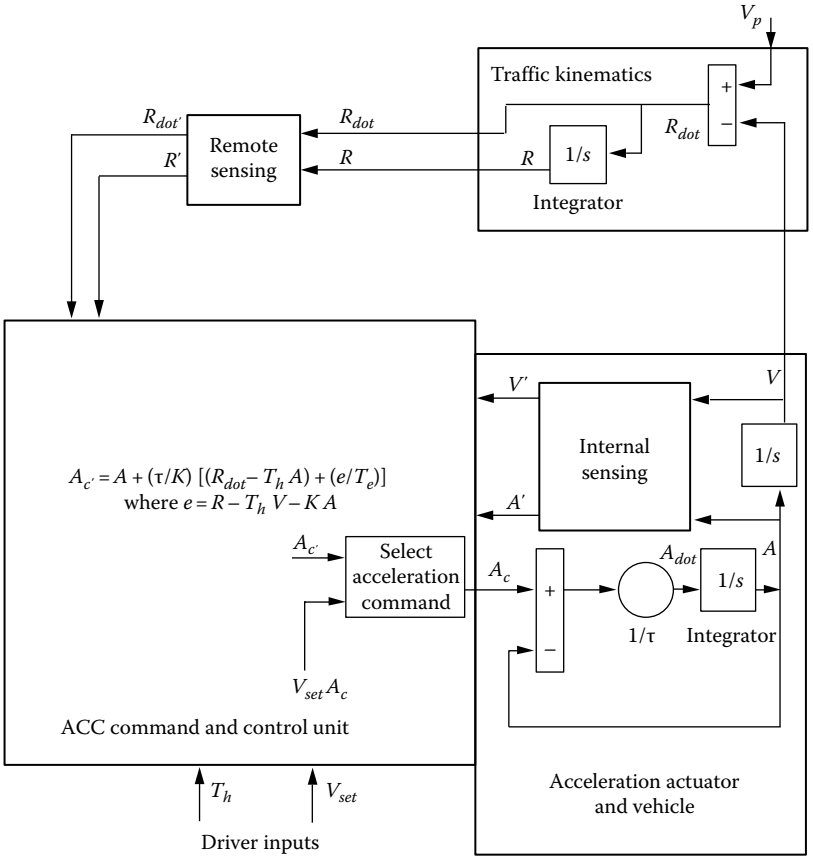


FIGURE 40.9 ACC using acceleration control.

the actuator if the design is to unify the actuator's characteristics into the overall system design. An error function satisfying these concepts is given by the following equation:

$$e = R - T_h V - K\ A \tag{40.18}$$

Step 2: The following equation can be used to define a hypothetical sliding surface that would converge toward $e = 0$ using an exponential approach:

$$Te\frac{de}{dt} + e = 0 \tag{40.19}$$

A point to observe is that *de/dt* contains *dA/dt*, which is A_{dot} in Equation 40.17. This means that we can solve Equations 40.17 through 40.19 simultaneously to determine the desired acceleration command A_c'. By doing this, we cancel the influences of τ and T_e on the open-loop transfer function between R_{dot} and *V*.

Step 3: The operations in step 2 given earlier are purely algebraic, although we are discussing differential equations. This is because we plan to use the integrators already existing in the traffic kinematics and the vehicle actuator and dynamics blocks to design a system that performs the desired ACC function for us. This way we avoid adding integrators, thereby keeping our conceptual design simpler and unifying the design. In this case, the state variables (which result from integrating) are *R, V,* and *A*. When the error *e* is approximately zero for an extended period of time, *R* is approximately equal to $T_h V$ as desired, and R_{dot} is approximately zero. By designing the ACC command and control unit in this manner, we have utilized the parts of the system so that the following relationship applies:

$$A_c' = A + \left(\frac{\tau}{K}\right)\left[R_{dot} - T_h A + \left(\frac{e}{T_e}\right)\right] \tag{40.20}$$

And the open-loop transfer function is as follows:

$$Y_o = \frac{V(s)}{R_{dot}(s)} = \frac{(T_e s + 1)}{\left[T_e K s^3 + (T_e T_h + K)s^2 + T_h s\right]} = \frac{1}{(Ks^2 + T_h s)} \tag{40.21}$$

Step 4: Due to the form of the equation for A_c', the parameter τ does not appear in the open-loop transfer function. And there is a pole/zero cancellation in Equation 40.21, which results from the form of the error dynamics specified by Equation 40.19. The result of this cancellation is that T_e does not appear in the final expression for the open-loop transfer function. Using this expression for Y_o, the closed-loop transfer function between V_p and *V* is as follows:

$$Y_c = \frac{V(s)}{V_p(s)} = \frac{1}{(Ks^2 + T_h s + 1)} \tag{40.22}$$

For a given value of T_h, the condition for the damping ratio to be greater than or equal to 1 is as follows:

$$K \le \left(\frac{T_h}{2}\right)^2 \tag{40.23}$$

Through this design process, we have theoretically arranged an ACC system that will meet our goals and will be free of overshoot as well. And furthermore, the parameter T_e is available for use in meeting other design challenges such as those related to driver ride-comfort issues [12] involving acceleration/deceleration adjustments.

For the magnitude of the transfer function (40.22) to be less than or equal to unity, the string stability condition is $K \leq T_h^2/2$, which is a less stringent restriction on K than that imposed by Equation 40.23. Equation 40.23 is based on eliminating overshoot in the time domain.

40.4.3 Unification Approach Applied to ACC with a Velocity Actuator

This case is similar to the one discussed for ACC using CCC (see Figure 40.7). Unification steps 1 through 4, which were discussed in the previous subsection, are outlined as follows for this case.

Step 1: The state variables are R and V. The functional purpose of this ACC system is to make $R = T_h V$. Hence, let $e = R - T_h V$. This form for e satisfies the functional purpose when it equals zero, and it contains all of the state variables as well.

Step 2: Then, consider $Te\ de/dt + e = 0$ as the desired approach to the sliding surface. This means that we want $V_{dot} = (1/T_h)\ [R_{dot} + (e/Te)]$. For an actuator and vehicle dynamics approximated by $\tau\ V_{dot} + V = V_c$, we want $V_c = V + (\tau/T_h)\ [R_{dot} + (e/Te)]$.

Step 3: Based upon the earlier expression for V_c, the open-loop transfer function is $Y_o = (1/T_h\ \mathrm{s})$. (This is the special case observed in the previous discussion of ACC using CCC.)

Step 4: The closed-loop transfer function is $(V_p(s)/V(s)) = Y_c = 1/(T_h\ s + 1)$. Clearly, this first-order system has no overshoot and it is string stable.

In summary, we have compensated for the actuator and vehicle dynamics while achieving constant time headway as the functional purpose that we chose. Interestingly, the value of the parameter T_e is available to use in tuning the system to have characteristics that are compatible with driver preferences and smoothing the ride performance of the ACC-equipped vehicle. However, the design depends upon having a good understanding of the value of τ and a pole/zero cancellation to remove T_e from the open- and closed-loop transfer functions. Also limits upon maximum acceleration and deceleration performance need to be considered in completing the overall design of the system.

40.4.4 ACC Systems with an Acceleration Actuator and No Available Acceleration Sensor

Although it is conceptually desirable to measure all of the state variables, there may be practical constraints on sensor availability or use such that the host vehicle's acceleration is not available for the ACC designer to use. In that case, the system is similar to that in Figure 40.9 except that A is not measured and the algorithm in the ACC command and control unit needs to be modified. For example, let $e = R - T_h V$ be the deviation from the desired sliding surface, where $e = 0$ if the system were on the sliding surface. Use $T_e\ de/dt + e = 0$ to solve for the desired value of acceleration command A_c. (Formally set V_{dot} equal to A_c.) The result is as follows:

$$A_c = \left(\frac{1}{T_h}\right)\left[R_{dot} + \left(\frac{R}{T_e}\right)\right]\frac{-V}{T_e} \tag{40.24}$$

Using the actuator equation $\tau\ dA/dt + A = A_c$, the open-loop transfer function is found to be

$$\frac{V(s)}{R_{dot}(s)} = \frac{(T_e s + 1)}{\left[T_e T_h \tau s^3 + T_e T_h s^2 + T_h s\right]} \tag{40.25}$$

And the closed-loop transfer function is as follows:

$$\frac{V(s)}{V_p(s)} = \frac{(T_e s + 1)}{\left[T_e T_h \tau s^3 + T_e T_h s^2 + (T_h + T_e)\, s + 1\right]} \tag{40.26}$$

With regard to string performance, the condition $|V(s)/V_p(s)|^2 \leq 1$ results in the following relationship:

$$\left[-T_e T_h \tau\, \omega^3 + (T_h + T_e)\omega\right]^2 + \left[1 - T_e T_h \omega^2\right]^2 - \left[(T_e \omega)^2 + 1\right] \geq 0 \tag{40.27}$$

As presented in Ref. [13], considerable algebraic effort leads to the conclusion that if $T_h \geq 2\tau$, then string stability can be achieved. From a design standpoint, this result has interesting implications. It means that, if the ACC designer and the actuator designer are not working together, the issue of string stability might be left unresolved as may be the case now for some current ACC systems. The issue is that the constraint $T_h \geq 2\tau$ involves quantities that have been regarded as set by considerations other than those associated with string performance. Given that T_h might be 1 s or less to satisfy driver preferences and $\tau \leq 0.5$ s may be difficult or impossible to achieve, there is a conflict. Perhaps, the actuator design might be modified to supply measured acceleration A to the communications bus, allowing A to be used in the ACC design and thereby avoiding the concern with making τ very small.

40.4.5 Summary of Related ACC Designs

Ref. [13] contains analytical work addressing a set of design cases for ACC systems. These cases cover command and control concepts based upon sliding surfaces (such as those already presented here), proportional control, and those nonautonomous systems that depend upon inter vehicle communication to obtain information on the acceleration $V_{p\,dot}$ (also symbolized as A_p) of the preceding vehicle. In order to summarize a number of similar design possibilities, the following generalized transfer function is introduced:

$$Y = \frac{N}{D} \quad \text{where} \quad N = n_2 s^2 + n_1 s + 1 \quad \text{and} \quad D = k_3 s^3 + k_2 s^2 + k_1 s + 1 \tag{40.28}$$

For systems of the form of Equation 40.28, the condition for string stability ($|Y|^2 \leq 1$), which is equivalent to $|D|^2 - |N|^2 \geq 0$, can be reduced to an expression of the following form:

$$\omega^2 (a + b\,\omega^2 + c\,\omega^4) \geq 0 \tag{40.29}$$

where

$$a = k_1{}^2 - 2k_2 + 2n_2 - n_1{}^2 \tag{40.30}$$

and

$$b = k_2{}^2 - k_1 k_3 - n_2{}^2 \tag{40.31}$$

and

$$c = k_3{}^2 \tag{40.32}$$

Inspection of (40.29) indicates that a sufficient condition for string stability is as follows:

$$a \geq 0 \quad \text{and} \quad b \geq 0 \tag{40.33}$$

(The quantity, c, is greater than or equal to zero as indicated by Equation 40.32.)

Table 40.1 presents four items, the special variables sensed or communicated, the actuator concept, the control command rule, and sufficient conditions for string stability in the sense of controlling overshoot. For all these ACC systems, the variables V, R, and R_{dot} (i.e., dR/dt) need to be sensed, but they are not listed as special variables in Table 40.1.

The last row of Table 40.1 represents a nonautonomous ACC system in which the special variable A_p is the acceleration of the preceding vehicle. In this case, it is presumed that A_p is communicated from the preceding vehicle to the ACC-equipped vehicle. The expressions given here represent a high level of abstraction, which means that component design and development at a lower level of abstraction may involve considerable technical challenge due to delays, noise, and inaccuracies in measuring and transmitting A_p.

The control command rule given in the last row of Table 40.1 is based upon a combination of first-order sliding surfaces that provide a second-order numerator for the transfer function, viz.,

let $e_1 = R - T_h V$, $e = T_1\, de_1/dt + e_1$, and $T_2\, de/dt + e = 0$. Based upon these choices, the combined result is as follows:

$$T_1 T_2 \left(\frac{d^2 e_1}{dt^2} \right) + (T_1 + T_2) de \frac{1}{dt} + e_1 = 0 \tag{40.34}$$

The expression for A_c in the last row of Table 40.1 is the result of employing the actuator equation and $e_1 = R - T_h V$ and Equation 40.34 to algebraically solve for A_c where τ_d has been included in case the value of τ is not accurately known. For $\tau/\tau_d = 1$ so that numerator terms cancel denominator terms, the open-loop transfer function Y_o becomes

$$Y_o = \frac{1}{T_h s} \quad \text{and} \quad Y = \frac{1}{(T_h s + 1)} \tag{40.35}$$

The string stability condition, $\tau/\tau_d \leq 1$, given in the last row of Table 40.1 is sufficient. Further treatment of this type of ACC system is given in [13,14]. Clearly, the values of the parameters T_1 and T_2 can be used to adjust for design objectives, uncertainties, and constraints in addition to those

TABLE 40.1
Summary of Related ACC Designs

Special Variables	Actuator Concept	Control Command Rule	For Overshoot Control
None	$T_c\, V_{dot} + V = V_c$	$V_c = (1/T_h)\,(R + T\, R_{dot})$	$T > T_c - T_h/2$
A	$\tau\, A_{dot} + A = A_c$	$Ac = (\tau_d/K)\,(e_{dot} + e/T_e) + A$ where $e_{dot} = de/dt$ and $e = R - T_h V - A$	$\tau/\tau_d \leq 1$ $K \leq T_h^2/2$
A, A_p	$\tau\, A_{dot} + A = A_c$	$Ac = [(\tau_d/T_1\, T_2\, T_h)\,(T_2\, T_1\, d^2R/dt^2) + (T_1 + T_2)\,(R_{dot} - T_h A) + R - T_h V)] + A$	$\tau/\tau_d \leq 1$

Note: R, R_{dot}, and V are sensed on board the vehicle. Also, $d^2R/dt^2 = A_p - A$.

discussed here. Systems of this type are being proposed as candidates for cooperative ACC systems and as steps toward automated highway systems (AHS) [13,14].

40.4.6 Evaluation of ACC Systems

As indicated in Figure 40.3, the same levels of abstraction as used in design may be applied in the evaluation of an existing system. Frequently, when evaluation is a trouble-shooting activity or purchasing related, one starts at the lowest level of abstraction by making sure that the system is packaged nicely, that it is installed well, and that there are no observable defects in any of the components. Tests may be specified to ensure that the components are functioning as designed and in accordance with acceptable performance objectives. International standards and technical-society standards frequently contain requirements concerning component performance. For example, [4,5] have provisions concerning sensors and command and control characteristics.

At the higher levels of abstraction, a limited amount of published work is available on evaluating how well an ACC system meets its functional purpose and the nature of the concepts employed in the system. With regard to functional purpose, a figure similar to Figure 40.10 has been used in [8,9] to compare driver-controlled following and ACC-controlled following. Figure 40.10 is based upon results from an ACC system using CCC as illustrated in Figure 40.6 and as described by Equations 40.8 through 40.10.

The results show that the ACC system achieved its purpose of reaching and holding $R_{dot} = 0$ and $R = T_h V$ very frequently when the system was operated by lay drivers in normal driving situations. In this sense, the system was shown to be robust in the face of the idiosyncrasies of normal traffic situations. And, as shown in Figure 40.10, the ACC system is clearly more predictable than the behavior of human drivers operating under the same types of road and driving conditions.

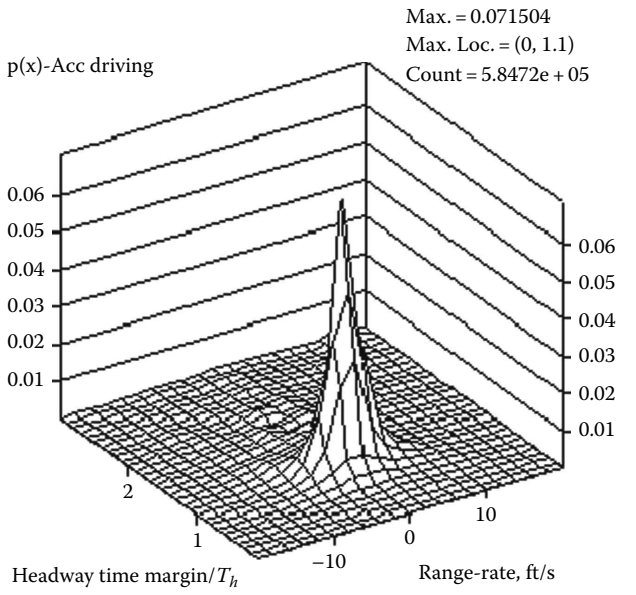
With regard to determining driver-centered ACC parameters that match the performance of drivers, the data used in constructing Figure 40.10 have been processed to indicate that the value of T used in Table 40.1 and Equation 40.9 should be chosen to be equal to approximately 10 s. This result of 10 s provides a good starting point for tuning the parameters T_e and $(T_1 + T_2)$ as listed in Table 40.1 and as used in Equations 40.19 and 40.34, respectively.

A noninvasive system-identification procedure has been devised for evaluating the operating characteristics of ACC systems in general [15]. In this procedure, the ACC-equipped vehicle and a confederate preceding vehicle are equipped with differential GPSs. Desired speeds and speed changes can be readily controlled if the preceding vehicle is equipped with CCC.

A series of evaluation test maneuvers are prescribed, and linear approximations to the command and control functions are determined in the neighborhood of various steady speed equilibrium points. The GPS data recorded are the positions and velocities of both vehicles and the GPS time at every 0.1 s during the test. The data recorded on board the preceding vehicle is synchronized with that recorded on the ACC vehicle using GPS time as recorded on both vehicles. The difference between the positions of the vehicles represents range R and the difference between velocities represents R_{dot}. The measurement system is noninvasive in the sense that no signals used in the ACC system are needed for the evaluation process.

For current ACC systems with $T_h \geq 1$ s, the accuracy requirements are not large with 1 m resolution in R being adequate for concept evaluation purposes. A time resolution of 0.1 s is achievable with readily available GPS equipment and this has been adequate.

The system-identification results for three current ACC systems indicate that a second-order linear system with R and V as state variables is adequate to represent perturbed performance. The input to the model is V_p, the velocity of the confederate leading vehicle. The identified values of the system parameters (the fitted parameters) are such that strings of these types of systems appear to be moderately unstable in the sense that a small change in speed of the lead vehicle will lead to a nonlinear disruption of string performance for a vehicle far enough back in the string. There would



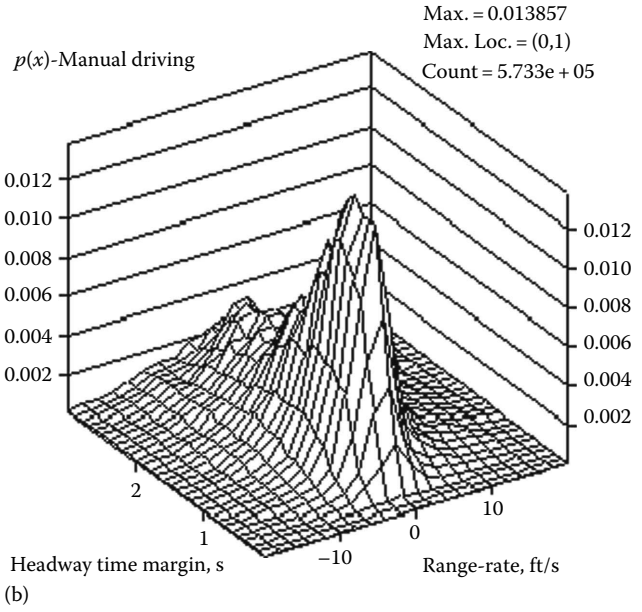


FIGURE 40.10 Evaluation of ACC functional purpose. Histograms of data for (a) ACC driving and (b) manual driving.

be a vehicle back in the string that needed to apply braking to avoid colliding with its preceding vehicle. This vehicle would need to break away from its control rule for following and adopt a modified rule for stopping or slowing if crashes are to be avoided. In this sense, the basic control structure of long strings of these types of vehicles breaks down due to safety concerns. The vehicles behind the stopped vehicle would need to stop, thereby leading to the possibility of stop-and-go driving. For example, predictions based on exercising the model indicate that the occurrence of four or more of these vehicles in a row would be likely to exhibit considerable overshoot for the last vehicle in response to moderate speed changes of the leading vehicle. Braking intervention by

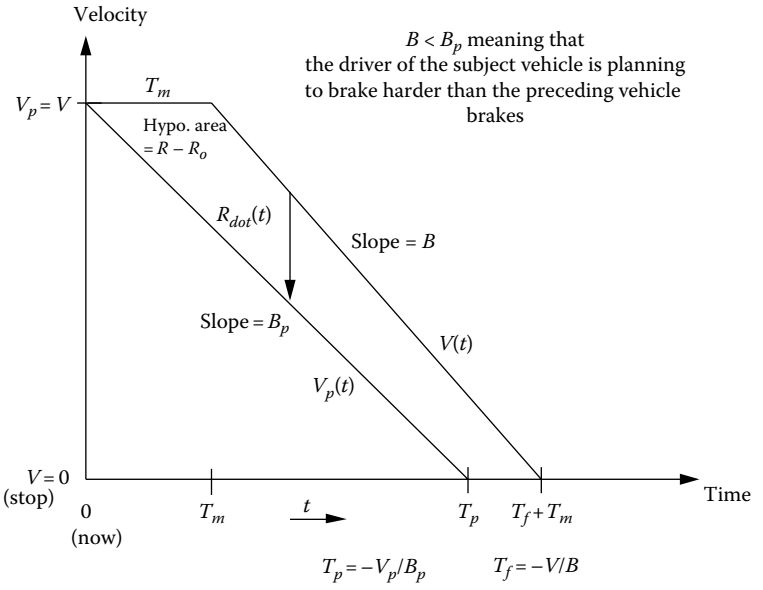
the driver of the last vehicle would appear to be likely [16] if the lead vehicle reduced speed by approximately 4 m/s in 10 s. More quantitative evaluations of this type are needed to generalize on the state of the art and the adequacy of current designs, particularly with regard to traffic flow implications and the practical importance of string stability issues as related to design concepts.

40.5 EMULATION OF THE LONGITUDINAL CONTROL PROPERTIES OF HUMAN DRIVING

There is a wealth of published material on models of manual driving, for example, see Refs. [17–19] and the papers referenced under driver-related topics in Section 40.9. (The driver's behavior while cornering is dealt with in Chapter 38.) The forms of these models depend upon the applications of the models as well as the concepts and theories that are familiar to the creators of the models. The data collected in recent operational test programs have included measurements of manual driving to use as a basis for comparison between driving with and without driver assistance systems [8,9]. This type of work appears to be in an exploratory phase. However, there is disagreement on the quality of the information available for predicting driver behavior. We are not going to classify driver models here. (See [17] for that.) Instead, we will discuss some recent work aimed at providing driver models to use in microscopic simulations of traffic flow including both manually driven and ACC-equipped vehicles.

The model presented here is based on ideas from Gipps [18] combined with the use of velocity time diagrams such as Figure 40.11 and parametric values chosen to match measured results from field tests involving manual driving [8,9]. Gipps's ideas involve the premise that drivers decide on a steady following distance based upon allowing a time (and/or a distance) margin that depends upon a safety-related estimate of what the preceding driver might do. Figure 40.11 has been created to aid in visualizing how ideas of this type might be viewed and analyzed.

Figure 40.11 shows a hypothetical situation in which the preceding vehicle decelerates at B_p m/s^2 and its velocity would reach zero at a time T_p. The conceptual idea behind Figure 40.11 is that



Hypo. area $= R - R_o = T_m V + T_f\ V/2 - T_p V_p/2$ or
$R = R_o + T_m V + G V^2$ where $G = ((1/2B_p) - (1/2B)) < 0$

FIGURE 40.11 Driver modeling concepts illustrated in a velocity–time diagram.

drivers base their future-driving behavior on the possibility that the hypothetical situation shown there might happen now. What the driver does amounts to allowing for holding the current speed V for T_m seconds and then decelerating at B m/s^2, which would lead to a stop at time $(T_m + T_f)$. The resulting range rate R_{dot} is the difference between V and V_p as indicated in Figure 40.11. The "hypo area" between the V and V_p curves is the integral of R_{dot}, which is the change in range R needed to conform to the hypothetical situation in the driver's mind.

The quantity "R_o" in the equations given in Figure 40.11 represents the clearance distance the driver chooses when both vehicles are stopped. The other terms in the equations result from applying kinematic relationships pertaining to constant deceleration stops. The point is that the driver's objective for steady following is as follows:

$$R = R_o + T_m V + G V^2 \text{ (as shown in Figure 40.11)} \tag{40.36}$$

Figure 40.11 was constructed for $V_p = V$, but similar figures could be constructed for V not equal to V_p. This would represent a transient situation such as closing or separating. Research on this modeling approach has not produced evidence indicating how drivers choose to approach steady following when (40.36) is used. However, there is information indicating that (40.36) is a better approximation to driver behavior in traffic than a constant time headway model [19]. Plots (two-dimensional histograms) of the frequency of various V and R combinations show that there is a well-defined ridge that approximately conforms to Equation 40.36. Attempts to use a number of currently popular microscopic models of traffic flow indicate that the Gipps type of model provides a better fit than several others that were tried [19].

Examination of driving behavior from field operational tests indicates that drivers have a wide range of proclivities for T_m. See, for example, Figure 40.12.

In addition, the parameter G varies considerably from driver to driver, but the values of G may be approximated from those for T_m by the following equation [20]:

$$G = -0.025 T_m + 0.013 \tag{40.37}$$

The implication of Equations 40.36 and 40.37 is that the time gap between vehicles will get smaller as speed increases. This may not be reassuring to safety advocates, but it may be viewed as positive regarding traffic flow on freeways.

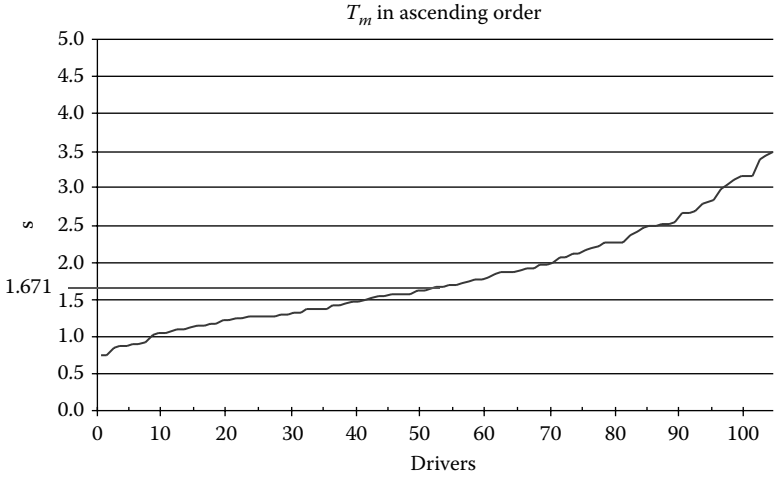


FIGURE 40.12 Linear coefficient in a nonlinear time gap relationship.

40.5.1 Nonlinear Time Gap ACC Systems

In the interest of reducing congestion and improving traffic volume, researchers are considering nonlinear time gap rules such as (40.36) and (40.37) for driver assistance systems. These could be seen as human centered for steady following, but since the transient aspects of human behavior may not be well enough understood, there does not appear to be a good grasp of how driver performance relates to string stability. The ability to emulate stop-and-go driving with microscopic driver models is still under development. Empirical fits to relatively few observations have been used to obtain models that have macroscopic verifications in selected circumstances described by fluid flow analogies.

Nevertheless, it is of interest to discuss the implications of choosing Equation 40.36 as the basis for an ACC system whose functional purpose is to operate in a human-centered manner as indicated by Equation 40.36. Although Equation 40.36 is nonlinear, the concepts and steps used in designing linear systems can be used. Example results for an ACC using CCC may be developed using $e = R - (R_o + T_m V + G V^2)$ as a starting point. Let $T_e\, e_{dot} + e = 0$ and $T_c\, V_{dot} + V = V_c$ for this case. Further, after developing the system equations as before and then obtaining the linear approximation to them at an equilibrium solution with $R = R_e$ and $V = V_e$, the open-loop transfer function is

$$Y_o = \frac{\Delta V}{\Delta R_{dot}} = \frac{(T_e s + 1)}{(ps)(T_e s\ +\ 1)} = \frac{1}{ps} \quad \left(\text{where } p = T_m + 2GV_e\right) \tag{40.38}$$

And the closed-loop transfer function is

$$Y_c = \frac{1}{(ps+1)} \tag{40.39}$$

This system will have string stability in the neighborhood of the equilibrium point at velocity V_e as long as $p = T_m + 2G\, V_e \geq 0$. To shorten the time gap as V_e increases, G needs to be negative, but this will lead to instability of the string structure when p becomes negative as speed is increased. The condition for $p = 0$ is $V_e = T_m/(-2\, G)$. For example, based upon the data discussed in the previous section, consider $T_m = 1.6$ s such that $G = -0.027$ and $V_e = 29.6$ m/s (66 mph). The point of this example is that instability of the string structure can start at a modest highway speed unless care is taken in selecting values for T and G. Suitable values of T and G that will be comfortable for drivers in ACC-equipped vehicles have not been determined as yet. And it appears that deviations from human-centered values might be acceptable for enhancing traffic flow.

The same type of finding applies when ACC using an acceleration-based actuator is employed. Perhaps, specially shaped nonlinear expressions for time gap, as a function of velocity, will be forthcoming. Perhaps switching rules can be employed to meet the special demands associated with both safety-related considerations and traffic flow requirements.

Another example of a nonlinear approach to ACC comes from considering a quantity known as "headway time margin (H_{tm})" where $H_{tm} = R/V$. As its name implies, H_{tm} provides an indication of the instantaneous time margin available to the driver or the controller in adjusting speed to achieve a desired range under the prevailing driving conditions. The original linear expression using $e = R - T_h\, V$ can be restated as follows using H_{tm}:

$$e = \left(\frac{R}{V}\right) - T_h \quad \text{where } H_{tm} = \frac{R}{V} \tag{40.40}$$

Although Equation 40.40 implies the same headway time relationship as $e = R - T_h V$, it represents a different form of algorithm for achieving steady following when we use $T_e\ de/dt + e = 0$ in formulating the control command for an ACC system. The result is as follows:

$$V_{dot} = \frac{V}{R}\left(R_{dot} + \frac{\left(R - T_h V\right)}{T_e} \right) \tag{40.41}$$

For a velocity actuator ($\tau\ V_{dot} + V = V_c$), Equation 40.41 would be substituted into the actuator equation to formulate the command control rule for evaluating V_c. Then (40.41) would be the differential equation describing the dynamic characteristics of the system. This description is similar to the one applying to the linear version except that $H_{tm} = R/V$ replaces T_h. This means that for situations in which H_{tm} is approximately equal to T_h, the nonlinear ACC system may perform much like the linear system listed in the first row of Table 40.1. The system will react aggressively when H_{tm} (i.e., R/V) is smaller than T_h and less aggressively when H_{tm} is larger than T_h. The idea is that the nonlinear system would be more aggressive in ensuring the desired time gap when the situation is more critical because H_{tm} is small.

40.6 FORWARD COLLISION WARNING

Although ACC has been developed commercially for a few years, forward collision warning (FCW) systems are emerging as a consumer product for the general public. Basic issues involve the reliability of these warnings as well as the number of false alarms that may be generated. There are a variety of concepts that have been proposed for implementing FCW systems. To a certain extent, the differences involve basic concepts concerning when and how to warn the driver. The following material provides conceptual examples of three types of warning systems that have slightly different functional purposes.

The first example is a human-centered system that is aimed at providing a warning of the possibility of a collision with a preceding vehicle or object when the driver enters a situation in which the driver is not braking but would be expected to be braking. This system is based on the idea of "time to impact, T_{ti}" where $T_{ti} = -(R/R_{dot})$ for $R_{dot} < 0$. The decision to warn the driver is based upon whether T_{ti} is less than a driver-selected threshold that is chosen by that driver to satisfy personal comfort limits. A typical value of this threshold is approximately 10 s, but drivers vary considerably in their preferences. This is a simple system to implement if the vehicle is already equipped for ACC and has sensors for measuring R and R_{dot}. The system may give frequent alarms [8], but the driver is expected to select a threshold that will meet that driver's comfort and acceptance tolerances. This type of system has been found to have promise in a limited number of trials, but the state of the art is such that researchers have not reported broadly based assessments of how effective this type of system would be.

A second example is based upon using a warning based upon anticipated deceleration needs. In this case, a threshold for the level of the deceleration generated in an ACC command controller could be used to command a warning signal to the driver. Since many ACC systems have a limit on the control authority allocated to the ACC functionality, the warning threshold might be set close to the control limit for the ACC system. For example, Equation 40.41 could be used with a threshold of −0.4 g. Analysis of identical braking situations has shown that this deceleration scheme may produce approximately one-third as many warnings as a time to impact system with $T_{ti} = 10$ s [8].

A third type of warning system can be based upon the range available as compared to the distance needed to stop before colliding. An idea of this type is currently being investigated in a field operational test. The alert algorithm employs a calculated miss distance to decide whether to warn [21].

Figure 40.13 has been constructed to illustrate ideas associated with the calculation of a desired range margin. This figure is like Figure 40.11 except that V_p is not equal to V. As shown in Figure 40.13, $R_{dot} = V_p - V$ has a negative value, which could indicate a crucial situation. The most demanding case is for $V_p(0) = 0$, i.e., a stopped preceding vehicle.

The situation shown in Figure 40.13 can provide the basis for several functional concepts for collision warning. For example, consider a design based upon the idea that the warned driver will decelerate at the same rate as the preceding vehicle in an emergency situation, that is, $B = B_w = B_p$ unless B_p is equal to zero. A warning is to be given if

$$R \le R_m + R_o \tag{40.42}$$

where R_o is the desired clearance at $V = 0$, and where

$$R_m = T_w V + \frac{\left(V_p^2 - V^2\right)}{2B_w} \tag{40.43}$$

If $B_p = 0$, a warning is to be given if the actual deceleration V_{dot} is not less than the value of B needed to stop in $R_m = R - R_o$, that is, $B = -V^2/2\,(R - R_o - T_w\,V)$.

The values of T_w and B_w are design choices. T_w corresponds to the time allowance provided to the driver, and B_w corresponds to the deceleration threshold at which an attentive driver is expected to have been braking already. A starting point for the designer to use in determining a suitable value for T_w might be 1.5 s, which is approximately the median value of T_m in Figure 40.12. Approximately −0.5 g could be a starting point for the designer to use in iterating through analyses, simulations, and experiments to find a suitable choice for the value of B_w.

Current issues involve finding values of T_w and B_w that provide a safety margin without too many false warnings. Experimental work [22] provides data indicating typical deceleration levels that drivers tend to use in situations requiring hard braking.

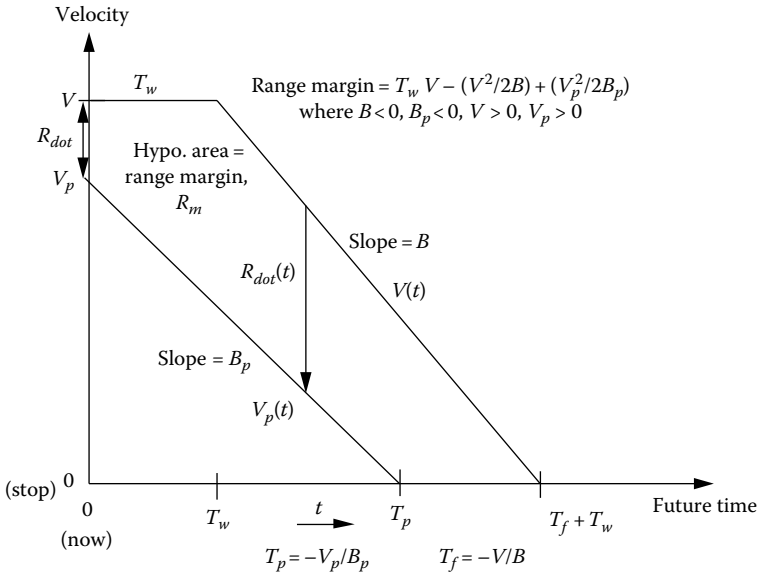


FIGURE 40.13 Range margin as used in FCW.

Another variation of the general theme portrayed by Figure 40.13 involves differentiating R_{dot} to obtain V_{pdot}, viz.,

$$V_{p\,dot} = V_{dot} + \frac{d\left(R_{dot}\right)}{dt} = B_p \tag{40.44}$$

Previously, we introduced a nonautonomous ACC system in which $V_{p\,dot}$ was communicated from the lead vehicle to the following vehicle. However, if R_{dot} could be differentiated practically and accurately enough, the vehicle-to-vehicle communication feature would not be needed and the ACC system could be autonomous. The use of Equation 40.44 for ACC systems has received little public attention, but in the case of FCW, the use of (40.44) to estimate B_p is being evaluated [21]. These possibilities open the way for unifying the design of new systems that combine the functions of both ACC and FCW.

Based on knowing B_p, Figure 40.13 could be interpreted differently. The current deceleration V_{dot} could be compared with a calculated deceleration B_n as needed to make R equal R_o at $V = 0$, viz., by examining Figure 40.13 and defining a quantity ΔR,

$$B_n = \frac{-V^2}{2\Delta R} \tag{40.45}$$

where

$$\left(R - R_o - TwV - \left(\frac{V_p^2}{2B_p}\right)\right) = \Delta R \tag{40.46}$$

If B_n is less than a threshold value for warning, the warning is given. For example, the threshold value might be chosen as −0.5 g. The point is that if the acceleration of the preceding vehicle is known (or estimated accurately enough), the collision warning system may have an advantage over an alert human driver who uses brake lights in sensing changes in the decelerations of preceding vehicles.

Nevertheless, there is a major concern with alarms that appear to be false to alert drivers. Such alarms may occur when the preceding vehicle is beginning to turn out of the path of the vehicle equipped with an FCW system or when an FCW-equipped following vehicle is beginning to pass a preceding vehicle. Perhaps advanced systems will someday use predictions of the path of the preceding vehicle to avoid apparently false alarms, but the meanings of these tactics are not currently clear in terms of tradeoffs between safety goals and driver acceptance.

40.7 AUTOMATED HIGHWAY SYSTEMS

The systems considered previously in this article depend upon the human driver's ability to supervise and interact with the driver assistance system in controlling the vehicle in a relatively continuous manner. In connection with ACC systems, the issue of the influence of ACC control characteristics on the structural performance of strings of vehicles was discussed with the intention of assuring that ACC systems might be designed in a manner that would not contribute to the tendency for stop-and-go driving to develop in congested driving situations. However, the headway time gap is an important factor in determining traffic flow at capacity. In the interest of safety, the time gaps employed in ACC systems may be slightly longer than those typically used by some drivers on congested freeways (see Chapter 39). This might reduce the density of traffic and thereby contribute to a reduction in traffic flow and capacity. Although advanced technology may

provide the impetus to go to shorter time gaps and clearance ranges, the driver may not be able to control the vehicle safely when quick actions are required. The delays in driver response and the bandwidth of driver control preclude the use of very short range clearances, which is the substance of high-capacity flows such as those anticipated for AHS. The behavior of ACC-equipped vehicles may provide insight into desired AHS features, and ACC systems may evolve toward systems using autopilot-like features rather than continuous driver supervision as the braking authority of ACC systems increases.

A literature review by Shladover [23] references and describes many of the ideas generated in early concepts of AHS structures. The use of platoons of closely spaced interacting vehicles is a common theme. The distance between platoons is designed to be long enough to allow the leader of a following platoon to be able to stop before colliding with the last vehicle in the preceding platoon. The goal is to increase traffic flow by having several vehicles at close range clearances in each platoon. The key technical challenge from a vehicle dynamics standpoint is to develop controls that will allow vehicles to travel safely and effectively at close range clearances.

In addition, there are challenges concerned with infrastructure requirements, institutional issues, and social and environmental issues. Nevertheless, congestion is sufficiently prevalent in certain locations around the world that considerable research effort has been supported by governmental transportation organizations.

Figure 40.14 is introduced to provide a simplified microscopic overview of steady traffic flow for vehicles whose speed V_{set} and steady-state time gap T_h are controlled automatically. In this example, all of the vehicles have the same control settings for V_{set} and T_h. In real traffic flows with driver-controlled vehicles, different drivers have different characteristics, which can be represented by distributions of V_{set} and T_h in microscopic models of traffic flow. These distributions lead to conflicts involving fast drivers wanting to pass slow drivers and whether a small or large gap is available for changing lanes. Clearly, the study of driver-controlled flow [24] is much more complicated than we are going to cover here in presenting basic ideas concerning ACC and AHS possibilities for reducing congestion.

Another point about Figure 40.14 is that it is a steady-state diagram, which means that transient disturbances are not shown. For string-stable ACC systems and AHS designs, the high-density end of the flow-density curve (beyond D_c and after the maximum flow F_c) is string stable such that disturbances may be expected to damp out exponentially. However, human drivers are not able to maintain stable strings in this region except for brief periods of synchronized flow that are disrupted by disturbances, often leading to stop-and-go driving [25].

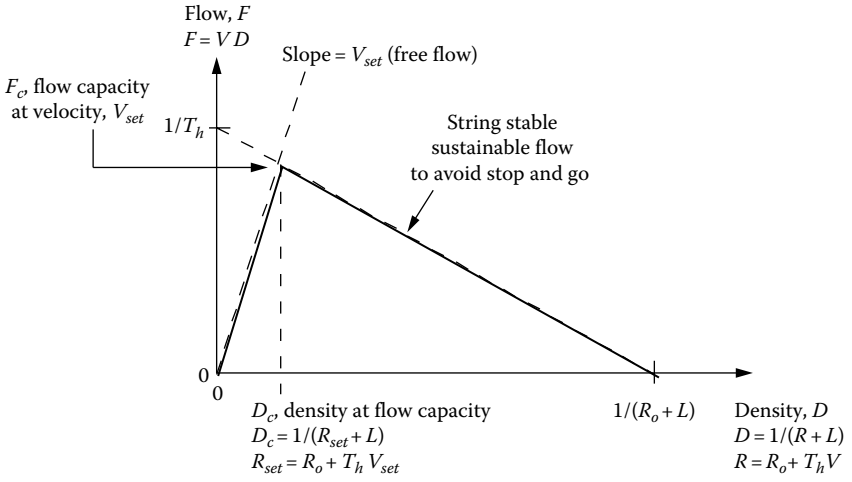


FIGURE 40.14 Steady-state flow with specified V_{set} and T_h.

Figure 40.14 essentially represents the steady-state behavior of identical vehicles equipped with identical string-stable ACC systems. The flow reduces aftercapacity because the ACC system enforces a desire time gap policy defined by $R = R_o + T_h V$. The quantity L represents the length of each vehicle, which is needed to consider density in terms of vehicles per meter and flow in terms of vehicles per second. Although flow may be a dynamic concept, the flow-density diagram is meant to provide insight to the traffic engineer in planning roads that will satisfy transportation needs. In this regard, the capacity F_c is a design goal and L has an important influence on the results.

Figure 40.15 is introduced to illustrate the importance of the platoon concept. For the purposes of this discussion, it is conceptually useful to think of a platoon as consisting of a fairly rigidly held configuration of several vehicles. In a sense, each platoon is like a packet of n vehicles, where n is the number of vehicles in the packet. If the control systems used in the platoon can hold the vehicles together at short clearance ranges, the length of the platoon will be short in terms of its length to number-of-vehicles ratio. The applicable equation for the length (L_p) of the platoon is as follows:

$$L_p = n\,L + (n-1)R_g \tag{40.47}$$

where R_g is the distance gap between vehicles.

Figure 40.15 shows three vehicles per platoon, but the number could be increased. One of the current research and design issues involves the number of vehicles that will perform suitably in a platoon. Nevertheless, the power of this approach concerning traffic flow can be readily demonstrated using three vehicle platoons. For example, consider a numerical example in which $L = 5$ m, $n = 3$, $R_g = 2.5$ m, $V_{set} = 30$ m/s, $R_o = 0$, and $T_h = 1$ s. The idea here is that each platoon leader is equipped with an ACC system that uses $V_{set} = 30$ m/s and $T_h = 1$. Lane A has a string of ACC-equipped vehicles and lane B has ACC vehicles leading three-vehicle platoons. Using the ideas portrayed in Figure 40.14, the flow at capacity for the lane A, with ACC vehicles and no platoons, is 0.857 vehicles/s (3085 vehicles/h). In contrast, the flow in lane B is 1.8 vehicles/s (6480 vehicles/h). These are both extraordinary maximum flow levels compared to what is sustained on freeways today (perhaps 2500 vehicles/h), but these are for idealized cases in which the influences of lane changing and entering and exiting are not considered. Nevertheless, the power of the platoon structure is evident.

If not already evident, the reason for the flow benefit of the platoon structure can be understood by considering packets. Each packet moves as a single entity with the length of the packet depending upon the number of vehicles in the packet. For the earlier example, the number of packets per meter in lane B is 1/(20 + 30) = 0.02 and the packet flow at capacity is (30)(0.02) = 0.6 packets/s. However, because there are three vehicles per packet, the vehicle flow is 1.8 vehicles/s. The point is that the packet structure effectively increases the density of the flow without consuming much additional space.

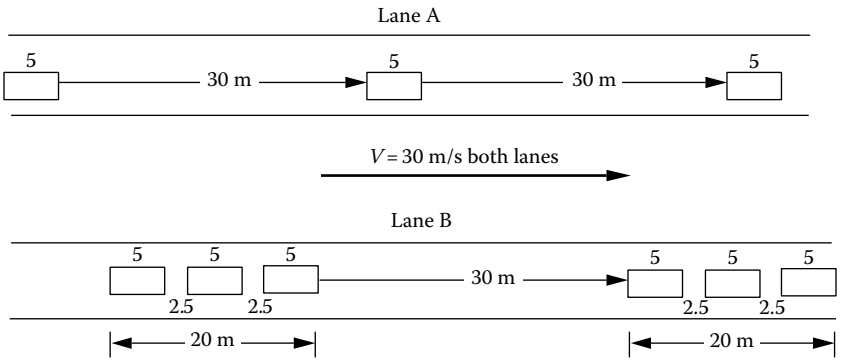


FIGURE 40.15 Example with three-vehicle platoons in lane B.

This density consideration is important in situations where current freeways are operated at over their capacity. Operating over capacity is brought about by the demand to get more cars on the freeway than the density at capacity provides. If the additional capacity brought about by platooning is not enough in an extreme situation, then the ACC lead vehicle could operate at a lower set speed (V_{set}) to accommodate the increased demand for density. This would reduce the flow, but it would increase the number of vehicles that could get onto the freeway and away from the congested area at times of peak congestion.

This same reasoning applies in the case of all ACC vehicles (lane A provided earlier). In the ACC case, the demand for flow is more likely to exceed the flow capacity required to meet the demand for density, that is, the demand for vehicles to get on the road. At periods of peak demand to get on the road, V_{set} may need to be reduced substantially to accommodate that demand. Nevertheless, the proposition is that, if the ACC systems are string stable in a structural sense, a steady flow may be sustained at a level beyond that achievable by manual driving under these high-density conditions. Research experiments are needed to demonstrate this in practice.

Up to this point in this discussion, the matter of the dynamics within a platoon has not been addressed. It has been presumed that the vehicles in the platoon can be controlled to maintain a relatively small distance between each other. The subject of string stability for platoons has been the subject of research concerning AHS. Currently, the work of Swaroop and Hedrick [26] is frequently referenced. This paper contains an elegantly presented mathematical set of theorems and methods indicating sufficient conditions for designing controllers that can achieve string stability for systems that maintain fixed distances between vehicles. A practical implication of implementing this approach is that it is necessary to measure and communicate the acceleration and speed of the lead vehicle to each following vehicle. For example, the third vehicle in a platoon would receive velocity and acceleration information from the lead vehicle in the platoon.

Although there are significant challenges in component and infrastructure design to be met, the existence of these theoretical results has provided the basis for justifying research and development of prototypes of these types of systems. Predictions involving even minimum two-vehicle platoons show significant contributions to increasing traffic flow and reducing congestion. For example, using the parametric choices for lane B as before except with $n = 2$, the maximum flow would be predicted to be 1.41 vehicles/s (5080 vehicles/h) with a density of 0.047 vehicles/m (47 vehicles/km). If the demand to get on the road implied a density of 60 vehicles/ km, V_{set} would need to be reduced to 20.8 m/s and the maximum flow would be 1.25 vehicles/s (4500 vehicles/h).

40.8 CLOSURE TO THE OVERVIEW OF BASIC CONCEPTS

The overview has used a design-oriented approach in describing functional features and performance possibilities for technologically advanced LC systems. The discussions have emphasized ACC, FCW, and AHS systems at the level of functional concepts and basic ideas. For the most part in this chapter, component design has been mentioned but not treated extensively. Rather, component properties have been treated as constraints on design possibilities for LC systems.

The success of these types of systems as consumer products is not currently known, but they are beginning to be offered in the marketplace. There will undoubtedly be many unanticipated issues to be understood and resolved as these systems are developed and introduced. Some of these issues may not have technological solutions. Changing the basic nature of the driving process is certain to have social, legal, and institutional ramifications.

The material that follows provides references to advanced treatments of topics pertaining to the dynamics of LC systems.

40.9 ADVANCED TOPICS

This section identifies a number of miscellaneous advanced topics related to the broad field of LC. The reader is referred to the indicated citations in the reference list for further detail and more in-depth discussions of each particular topic. By and large, the vast majority of these references are from the last 15 years or so. More seasoned works have been selected to help address some of the more traditional LC topics. Many of these topics have had a waning research interest in recent years or have been more competently examined, in large measure, by those earlier studies.

40.9.1 Driver Closed-Loop Braking Performance under Limit Conditions

Examples of driver closed-loop braking performance in emergency-like or hard-braking scenarios are provided in such studies as Rompe [27,28], McLean [29], Newcomb [30], and more recently by Fambro [31]. Rompe and Fambro provide experimental measurements of drivers in passenger cars and light trucks. McLean and Newcomb provide not only similar experimental measurements of passenger car and commercial vehicle drivers during heavy braking but also a fairly comprehensive and validated model of the human driver engaged in such stopping maneuvers. Hisaoka [32] provides a more recent study of measuring and modeling driver behavior under braking-in-a-turn conditions. Various aspects of understanding and modeling basic human control behavior over a range of operating conditions are also discussed in MacAdam [17].

40.9.2 Driver Headway Control Behavior

The difficult task of studying and characterizing driver LC behavior under more routine or less demanding longitudinal conditions, such as headway control, is also examined in a number of other works. In many of these cases, driver LC is often intermittent, due to competing distractions or low attention demands imposed on the driver by the control task. Examples would include the works of Fancher [33,34], Eyre [35], Gerdes [36], Girault [37], Lee [38], Lu [39], Mayr [40], Peng [24], Stankovic [41], and MacAdam [42]. Most of these works include measurements of driver behavior as well as efforts to model such behavior. In some cases, interaction with driver-assist systems is also a major component of the study such as in Bengtsson [43] or Fancher [33].

40.9.3 Sliding Control

Example studies that have employed sliding control techniques to represent longitudinal controllers or driver assists are provided in such works as Gerdes [44], Hedrick [45], Lee [38,46], Lim [47], Lu [39], and No [48].

40.9.4 Driver Longitudinal Control Behavior When Entering Curves

The braking behavior of drivers, when approaching or entering curves, is examined in a number of studies. Example works that have measured and modeled this behavior are provided by Levison [49,50], Savkoor [51], and Yoshimoto [52].

40.9.5 Nonlinear Control

Examples of nonlinear LC applications include Fritz [53], Gerdes [44], Hedrick [45], Kalkkuhl [54], Lee [46], Lim [47], Omae [55], Schiehlen [56], Seshagiri [57], Warnick [58], and Yanakiev [59].

40.9.6 Control Hysteresis and Saturation

Issues related to the nonlinear topics of control hysteresis and control saturation are addressed in such studies as Gerdes [44], Girault [37], and Warnick [58].

40.9.7 Control Delays

The importance of transport delays on the performance of LC systems is given in such works as Huang [60,61], Liu [62], Wang [63], and Yanakiev [59].

40.9.8 Platoons and String Stability

Papers related to the topics of platoons or longitudinal string stability are found in works such as Eyre [35], Girault [37], Hedrick [10], Huang [61], Lee [38,46], Liu [62], Lu [64], No [48], Nobe [65], Omae [55], Rajamani [66], Schiehlen [56], Seshagiri [57], Stankovic [41], Swaroop [67], Tsugawa [68], Wang [69], Warnick [58], and Yanakiev [59].

40.9.9 Artificial Intelligence Methods Applied to Longitudinal Control

Neural network and fuzzy logic approaches to analyzing and modeling the LC behavior of both drivers and advanced vehicle control systems are seen in such works as Kalkkuhl [54], Lee [38,46], MacAdam [42,70], Nobe [65], and Wang [63].

REFERENCES

1. Rasmussen, J., Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, 13(3), 257–266, 1983.
2. Slotine, J. and Li, W., *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice Hall, 1991.
3 Ulsoy, A. G., *Vehicle Control Systems*. Ann Arbor, MI: College of Engineering and the Transportation Research Institute (UMTRI), University of Michigan, 1992.
4. Transport Information and Control Systems—Adaptive Cruise Control Systems—Performance Requirements and Test Procedures, ISO Standard No. 15622: 2010, ISO, 2010.
5. Adaptive Cruise Control (ACC) Human Factors: Operating Characteristics and User Interface—Standard, Safety and Human Factors Standards Committee, Standard J2399, SAE, Dec. 15, 2003.
6. Riley, B., Kuo, G.,Schwartz, B., Zumberge, J., and Shipp, K., Development of a controlled braking strategy for vehicle Adaptive Cruise Control, *Brake Technology ABS/TCS Systems, NVH, and Foundation Brakes*, SAE SP-1537, 2000.
7. Winner, H. and Olbrich, H., Major design parameters of Adaptive Cruise Control, *Proceedings of the International Symposium on Advanced Vehicle Control (AVEC)*, Nagoya, Japan, 1998.
8. Fancher, P. et al., Fostering the development, evaluation, and deployment of forward crash-avoidance systems (Focas), UMTRI, Tech. Rept. UMTRI-2000-27, 2000.
9. Fancher, P. et al., Intelligent Cruise Control (ICC) Field Operational Test, UMTRI, Tech. Rept. DOT-HS-808-849, 1998.
10. Hedrick, J. K., Practical string stability. *18th IAVSD Symposium of the Dynamics of Vehicles on Roads and on Tracks*, Atsugi, Kanagawa, Japan, 2003.
11. Ioannou, P. and Chien, C. C., Autonomous Intelligent Cruise Control. *IEEE Transactions on Vehicular Technology*, 42(5), 671–677, 1993.
12. Yamamuma, Y. et al., An ACC design method for achieving both string stability and ride comfort. *Proceedings of AVEC'02*, Hiroshima, Japan, 2002.
13. Zhou, J. and Peng, H., String stability conditions of Adaptive Cruise Control algorithms. *IFAC Proceedings*, Washington, DC, 2003.
14. Rajamani, R. and Zhu, C., Semi-autonomous Adaptive Cruise Control Systems. *IEEE Transactions on Vehicular Technology*, 51(5), 1186–1192, 2002.
15. Bareket, Z. et al., Methodology for assessing Adaptive Cruise Control behavior. *IEEE Transactions on ITS (Special Issue on Automatic Cruise Control)*, 4(3), 123–131, 2003.
16. Fancher, P., Research on desirable Adaptive Cruise Control behavior in traffic streams, UMTRI, Tech. Rept. UMTRI-2002-16, 2002.
17. MacAdam, C. C., Understanding and modeling the human driver. *Vehicle System Dynamics*, 40(1–3), 101–134, 2003.
18. Gipps, P. G., A behavioral car-following model for computer simulation. *Transportation Research-B*, 15(B), 105–111, 1981.

19. Lee, K. W. and Peng, H., Identification and verification of a longitudinal human driving model for collision warning and avoidance systems. *International Journal of Vehicle Autonomous Systems*, Vol. 2, Nos. 1/2, 2004.
20. Fancher, P. et al., Research on desirable Adaptive Cruise Control behavior in traffic streams, UMTRI, Tech. Rept. UMTRI-2003-14 (Phase 2), 2003.
21. Brunson, S., Alert algorithm development program—NHTSA rear-end collision alert algorithm, U.S. DOT, Tech. Rept. DOT HS 809 526, 2002.
22. LeBlanc, D. et al., Forward collision warning: Preliminary requirements for crash alert timing—Paper 2001-01-0462. *SAE Proceedings*, Detroit, MI, 2001.
23. Shladover, S., Review of the state of development of Advanced Vehicle Control Systems (AVCS). *Vehicle System Dynamics*, 24, 551–595, 1995.
24. Peng, H., Evaluation of driver assistance systems—A human centered approach. *AVEC Proceedings*, Hiroshima, Japan, 2002.
25. Kerner, B., Congested traffic flow: Observations and theory. *Proceedings of the Transportation Research Board 78th Annual Meeting*, Washington, DC, 1999.
26. Swaroop, D. and Hedrick, J. K., String stability of interconnected systems. *IEEE Transactions on Automatic Control*, 41(3), 349–357, 1996.
27. Rompe, K., Schindler, A., and Wallrich, M., Comparison of the braking performance achieved by average drivers in vehicles with standard and anti wheel lock brake systems. *SAE International Congress and Exposition*, SAE, Detroit, MI, 1987.
28. Rompe, K., Wallrich, M., and Schindler, A., Advantages of anti wheel lock system (Abs) for the average driver in difficult driving situations. *Proceedings of the 16th International Symposium on Automotive Technology and Automation*, University of Florence, Province Florence, Tuscany, Italy, 1987.
29. McLean, D., Newcomb, T. P., and Spurr, R. T., Simulation of driver behavior during braking. *IMechE Conference on Braking of Road Vehicles*, IME, Loughborough, U.K., 1976.
30. Newcomb, T. P., Driver behavior during braking. *Proceedings of the SAE West Coast International Meeting*, SAE, Seattle, WA, 1981.
31. Fambro, D. B. et al., Driver braking performance in stopping sight distance situations. *Transportation Research Record*, 1701, 9–16, 2000.
32. Hisaoka, Y., Yamamoto, M., and Okada, A., Closed-loop analysis of vehicle behavior during braking in a turn. *JSAE Review*, 20(4), 537–542, 1999.
33. Fancher, P. S. and Bareket, Z., Evolving model for studying driver-vehicle system performance in longitudinal control of headway. *Transportation Research Record*, 1631, 13–19, 1998.
34. Fancher, P. S. and Bareket, Z., Manual and automatic control of the longitudinal dynamics of individual motor vehicles. *Vehicle System Dynamics*, 33(Suppl. 2000), 270–281, 2000.
35. Eyre, J., Yanakiev, D., and Kanellakopoulos, I., Simplified framework for string stability analysis of automated vehicles. *Vehicle System Dynamics*, 30(5), 375–405, 1988.
36. Gerdes, J. C., Rossetter, E. J., and Saur, U., Combining lanekeeping and vehicle following with hazard maps. *Vehicle System Dynamics*, 36, 4–5, 2001.
37. Girault, A. and Yovine, S., Stability analysis of a longitudinal control law for autonomous vehicles. *Proceedings of the IEEE Conference on Decision and Control*, Phoenix, AZ, 1999.
38. Lee, G. D. and Kim, S. W., A longitudinal control system for a platoon of vehicles using a fuzzy-sliding mode algorithm. *Mechatronics*, 12, 97–118, 2002.
39. Lu, X.-Y., Hedrick, J. K., and Drew, M., ACC/Cacc—Control design, stability and robust performance. *Proceedings of the American Control Conference*, Anchorage, AL, 2002.
40. Mayr, R. and Bauer, O., Safety issues in Intelligent Cruise Control. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Boulder, CO, 1999.
41. Stankovic, S. S., Mladenovic, S. M., and Siljak, D. D., Headway control of a platoon of vehicles: Inclusion principle and LQ optimization. *Proceedings of the IEEE Conference on Decision and Control*, Tampa, FL, 1998.
42. MacAdam, C. C. et al., Using neural networks to identify driving style and headway control behaviour of drivers. *Proceeding of the Dynamics of Vehicles on Roads and on Tracks*, Budapest, Hungary, 1997.
43. Bengtsson, J., Johansson, R., and Sjogren, A., Modeling of drivers' longitudinal behavior. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, Como, Italy, 2001.
44. Gerdes, J. C. and Hedrick, J. K., Hysteresis control of nonlinear single-acting actuators as applied to brake/throttle switching. *Proceedings of the American Control Conference*, San Diego, CA, 1999.
45. Hedrick, J. K., Nonlinear controller design for automated vehicle applications. *IEEE Conference Publication*, UKACC International Conference on Control, (Control '98), vol. 1, pp. 23–32, Swansea, Wales, UK, Sept. 1–4, 1988.

46. Lee, H. and Tomizuka, M., Adaptive vehicle traction force control for Intelligent Vehicle Highway Systems (IVHS). *IEEE Transactions on Industrial Electronics*, 50, 37–47, 2003.
47. Lim, E. H. M. and Hedrick, J. K., Lateral and longitudinal vehicle control coupling for automated vehicle operation. *Proceedings of the American Control Conference*, San Diego, CA, 1999.
48. No, T. S., Chong, K.-T., and Roh, D.-H., A Lyapunov function approach to longitudinal control of vehicles in a platoon. *IEEE Transactions on Vehicular Technology*, 50, 116–125, 2001.
49. Levison, W. H. et al., Modification and partial validation of the driver/vehicle module. *Transportation Research Record*, 1803, 52–58, 2002.
50. Levison, W. H. et al., Stopping-distance model for driver speed decision making in curve approach. *Proceedings of the Human Factors and Ergonomics Society*, 2, 1222–1226, 1998.
51. Savkoor, A. R. and Ausejo, S., Analysis of driver's steering and speed control strategies in curve negotiation. *Vehicle System Dynamics*, 33(Suppl. 2000), 94–109, 2000.
52. Yoshimoto, K., Katoh, M., and Inoue, K., A vision-based speed control algorithm for autonomous driving. *Proceedings of AVEC 2000 5th International Symposium on Advanced Vehicle Control*, Ann Arbor, MI, 2000.
53. Fritz, H., Longitudinal and lateral control of heavy duty trucks for automated vehicle following in mixed traffic: Experimental results from the chauffeur project. *IEEE Conference on Control Applications Proceedings*, Kona, HI, 1999.
54. Kalkkuhl, J., Hunt, K. J., and Fritz, H., Fem-based neural-network approach to nonlinear modeling with application to longitudinal vehicle dynamics control. *IEEE Transactions on Neural Networks*, 10(4), 885–897, 1999.
55. Omae, M. and Fujioka, T., Longitudinal control of platoon treated as one dynamic system. *Dynamic Systems and Control Division American Society of Mechanical Engineers, Dynamic Systems and Control Division*, 64, 691–703, 1998.
56. Schiehlen, W., Motion control of vehicles in convoy. *International Conference on Control of Oscillations and Chaos, Proceedings*, St. Petersburg, Russia, 2000.
57. Seshagiri, S. and Khalil, H. K., Longitudinal adaptive control of a platoon of vehicles. *Proceedings of the American Control Conference*, San Diego, CA, 1999.
58. Warnick, S. C. and Rodriguez, A. A., Systematic antiwindup strategy and the longitudinal control of a platoon of vehicles with control saturations. *IEEE Transactions on Vehicular Technology*, 49(3), 1006–1016, 2000.
59. Yanakiev, D. and Kanellakopoulos, I., Longitudinal control of automated CHVs with significant actuator delays. *IEEE Transactions on Vehicular Technology*, 50, 1289–1297, 2001.
60. Huang, S. and Ren, W., Vehicle longitudinal control using throttles and brakes. *Robotics and Autonomous Systems*, 26(4), 241–253, 1999.
61. Huang, S. and Ren, W., Longitudinal control with time delay in platooning. *IEEE Proceedings: Control Theory and Applications*, Vol. 145, Issue 2, pp. 211–217, March, 1998.
62. Liu, X. et al., Effects of communication delay on string stability in vehicle platoons. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Singapore, 2001.
63. Wang, J., Chao, S. Y., and Agogino, A. M., Validation and fusion of longitudinal positioning sensors in AVCs. *Proceedings of the American Control Conference*, San Diego, CA, 1999.
64. Lu, X.-Y. and Hedrick, K. J., Longitudinal control algorithm for automated vehicle merging. *Proceedings of the IEEE Conference on Decision and Control*, Sydney, New South Wales, Australia, 2000.
65. Nobe, S. A. and Wang, F.-Y., An overview of recent developments in automated lateral and longitudinal vehicle controls. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Tucson, AZ, 2001.
66. Rajamani, R. et al., Design and experimental implementation of control for a platoon of automated vehicles. *Dynamic Systems and Control Division American Society of Mechanical Engineers, Dynamic Systems and Control Division*, 64, 681–689, 1998.
67. Swaroop, D., Hedrick, J. K., and Choi, S. B., Direct adaptive longitudinal control of vehicle platoons. *IEEE Transactions on Vehicular Technology*, 50, 150–161, 2001.
68. Tsugawa, S. et al., A cooperative driving system with automated vehicles and inter-vehicle communications in Demo 2000. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Singapore, 2001.
69. Wang, J., Chao, S. Y., and Agogino, A. M., Sensor noise model development of a longitudinal positioning system for AVCs. *Proceedings of the American Control Conference*, San Diego, CA, 1999.
70. MacAdam, C. C. and Bareket, Z., Adaptive neural network characterizations of driver longitudinal control behavior. *Proceedings of the International Symposium on Advanced Vehicle Control—AVEC*, Nagoya, Japan, 1998.

Part VIII

Road Accident Reconstruction and Passive Safety

41 Analysis and Reconstruction of Road Accidents

Horst Ecker

CONTENTS

41.1 Introduction 1584
41.1.1 Accident Analysis: Who and Why 1584
41.1.2 Information Sources 1585
41.2 Accident Scenarios 1585
41.3 Sequence of Accident Phases 1586
41.3.1 Precrash Phase 1586
41.3.1.1 Calculation of a Basic Braking Maneuver 1586
41.3.1.2 Estimation of Braking Deceleration 1587
41.3.1.3 Analysis of Evidence at the Accident Scene 1589
41.3.2 Crash Phase 1589
41.3.3 Postcrash Phase 1591
41.3.4 Reconstruction Strategies 1592
41.4 Models for Accident Analysis 1593
41.4.1 Vehicle Models 1593
41.4.1.1 Kinematic Models 1593
41.4.1.2 Kinetic Models 1594
41.4.2 Tire Force Models 1594
41.4.2.1 "Magic Formula" Tire Force Model 1595
41.4.2.2 VCRware Tire Force Model 1595
41.4.2.3 PC Crash Linear Tire Force Model 1595
41.4.2.4 TMEasy Tire Force Model 1596
41.4.2.5 Other Tire Force Models 1596
41.4.2.6 Evaluation and Comparison of Tire Force Models 1596
41.4.3 Crash Models 1597
41.4.3.1 Planar Impact Mechanics 1597
41.4.3.2 Energy-Based Methods 1600
41.4.3.3 Alternative Methods 1601
41.5 Software Tools for Accident Analysis 1602
41.5.1 HVE and HVE-2D 1602
41.5.1.1 Program Modules Compatible with HVE 1603
41.5.1.2 Advanced Features in HVE 1604
41.5.2 PC Crash 1605
41.5.3 VCRware 1606
41.5.4 Other Software 1607
41.6 Special Topics in Accident Analysis 1607
41.6.1 Motorcycle Accidents 1607
41.6.2 Accidents Involving Pedestrians 1608
References 1608

41.1 INTRODUCTION

The story of the evolution of motorized vehicles is also the story of vehicle accidents. Nicolas Cugnot, a French military engineer, designed the very first self-propelled car, a steam-powered road vehicle, in 1769. In the following year, he built another steam-powered tricycle that carried four passengers. Only a year later, Cugnot drove one of his road vehicles into a stone wall, which would make him the first known person to get in a motor vehicle accident. Although it is disputed whether this accident ever happened, it is quite likely that developing new technology at this time would cause some unforeseen events and accidents.

According to [1], the first recorded automobile fatality occurred in a small town in the Irish Midlands in 1869. Mary Ward, a celebrated microscopist, artist, astronomer, and naturalist, fell from a steam carriage and died of severe injuries from its heavy iron wheels.

If the accident mentioned earlier actually happened to Mr. Cugnot, then the claim [2] that the world's first automobile accident occurred in Ohio City/Ohio in 1891 must probably be confined to gasoline-powered vehicles in the United States. James W. Lambert, a very successful inventor holding more than 600 patents mostly affiliated with the automobile industry, was driving in his vehicle. When a wheel hit a tree root, the car careened out of control and smashed into a hitching post. Luckily, injuries from this accident were minor.

A more severe accident with less luck for the vehicle occupants happened a few years later, on February 25, 1899, in Great Britain. Edwin Sewell and Major Richer are the first recorded driver and passenger fatalities, respectively. They were thrown from their vehicle on Grove Hill Road in Harrow on the Hill, Middlesex, London. In remembrance of this tragic event, a roadside plaque was unveiled at the accident scene seventy years later [3].

Since these early days, road traffic has grown immensely all over the world. Along with this growth, also traffic accidents have reached a frequency that is hard to believe and has dramatic consequences. According to a study by the WHO for Europe, every year, road traffic accidents kill about 127,000 people, injure at least some 2.4 million, and kill more children and young people aged 5–29 than any other causes of death. Worldwide, road incidents result in the death of an estimated 1.3 million people each year and between 20 and 50 million sustain nonfatal injuries [4].

Although such numbers should shock people, they are too difficult to imagine, too surreal, and therefore don't really impress the public. A traffic accident moves people if they bear relation to the location, the vehicles, or the victims. If a relative, a friend, or a celebrity is involved in an incident, then people are willing to focus on what is happening every day, every minute, somewhere on streets and roads.

Vehicle accidents involving celebrities also put the work of accident reconstructionists in the center of focus. Movie hero James Dean was killed in 1955 in a Porsche 550 Spyder in an almost head-on collision with a 1950 Ford Tudor. Millions of fans of the actor mourned, but also the cause of the accident was debated and investigated several times. More recently, the tragic vehicle accident on August 31, 1997, in Paris that caused the death of Lady Diana, Princess of Wales, moved even more people all around the world. Since the circumstances of this fatal crash have been quite unclear, accident reconstruction did play a major role in this case [5] and probably made this car accident the most thoroughly and comprehensively investigated one so far.

41.1.1 Accident Analysis: Who and Why

The analysis and reconstruction of road vehicle accidents are conducted by various institutions within a country and also differently organized in different countries. Most often, police traffic units are the first to be involved in the process. Typically, they will secure the accident scene, collect and document evidence, take vehicles into custody, etc. Sometimes accident reconstruction experts are involved at this early stage, and sometimes they act at a later stage. These experts may be specially trained officers within police units, and they may be employed by other government

organizations or by law offices, professors within academic institutions, or members of forensic offices specializing in accident reconstruction. Vehicle manufacturers also have their own departments with specialists in the field of accident analysis and reconstruction.

The obvious reason for analyzing a vehicle traffic accident is to determine the cause of the accident. If a criminal act was the cause, a criminal procedure will be started against the individual(s) involved, and if held criminally liable, those individuals may be subject to criminal penalties. Even if no crime occurred, lawsuits frequently result from vehicle accidents. This can be true whether the accident was caused by human error, defective design, or construction of the vehicles involved or even defects in the design or construction of the road or the signage in the vicinity of the accident location.

At least since the publication of the famous book "Unsafe at Any Speed" [6], product liability has become an important issue in the automotive industry. Manufacturers are often accused of having produced defective products that either caused or significantly contributed to accidents. Defective design cases can be especially challenging for the accident reconstructionist.

41.1.2 Information Sources

Before the topic of this chapter is discussed in detail, a brief overview of major literature resources is given. During the last decades, significant research efforts were made at various institutions and in different countries, and meanwhile a wealth of information is available. Comprehensive material on vehicle accident reconstruction and numerous of references can be found in these up-to-date books [7–11] and the following journals:

- *Accident Analysis & Prevention*, Elsevier, ISSN: 0001-4575
- *The Accident Reconstruction Journal* (*AR Journal*), ISSN 1057-8153
- *VKU Verkehrsunfall und Fahrzeugtechnik*, Vieweg+Teubner, ISSN: 0724-2050

Note that these are by no means complete lists and shall serve only as a base to start from.

41.2 ACCIDENT SCENARIOS

It is impossible to develop and present a unified approach to the analysis of vehicle accidents. The reason for this is the fact that too many different scenarios of accidents are possible and do actually occur in daily life. Let us start with a list of different motorized vehicles, which populate roads and highways:

- Four-wheeled vehicles (cars, ATVs, light trucks, etc.)
- Powered two-wheelers (PWTs) (motorcycles, mopeds, etc.)
- Multiwheeled vehicles (trucks, busses, etc.)
- Multivehicle combinations (tractor–trailer, RV with trailing car, etc.)
- Special vehicles (trikes, articulated vehicles, off-road vehicles, etc.)

Although the laws of physics hold for all these vehicles, an accident may appear very different if, for example, either a passenger car or a motorcycle is involved. Since each group of vehicles has its own specific properties, accident analysis needs to consider such particular features. Consequently, there are comprehensive publications on accident reconstruction devoted to certain groups of vehicles like motorcycles [12], trucks [13], or busses [14].

Not only do we have to consider the various aforementioned motorized vehicles, but collisions may also occur in any combination with

- Other motorized vehicles
- Nonmotorized vehicles (bicycle, tricycle, etc.)
- Pedestrians

- Animals
- Nonmoving objects (tree, mast, building, guardrails, etc.)
- The road or a part of the environment (curb, slope, river, lake, ditch, etc.)

From this compilation, it is easy to imagine that there are many combinations possible and almost each of them has its subtle peculiarity. There are, of course, standard situations like a pedestrian hit by a car on the crosswalk, a car versus car collision at an intersection, or a motorcycle going off the road in a turn that can be analyzed by routine. But quite often, special details have to be taken into account that make an accident unique and difficult to analyze.

41.3 SEQUENCE OF ACCIDENT PHASES

There is a distinctive pattern that is common to practically all accidents. The event can be subdivided into three different phases: precrash, collision(s), and postcrash. Depending on the type of accident, the kind of vehicles, or nonvehicles involved, each individual phase may be different, but the sequence will be the same and can be used to structure the entire event.

In the following sections, the characteristics of all three phases will be discussed. To illustrate and underline the principle issues of each phase, the simple case of a straight-line rear end collision is used as an example. More complex scenarios will be discussed later in this chapter.

41.3.1 Precrash Phase

The precrash phase is defined as the period of time before the first and unintended contact with another vehicle, human being, animal, or object. In general, this phase starts with a stimulus for the driver of a vehicle, which triggers a sequence of visual information processing. This process involves the perception of a hazard or threat, the decision-making, and the response to avoid the accident. The beginning of this sequence is largely dominated by the cognitive abilities of the driver [15,16]. To analyze an accident, especially when environmental factors (like darkness and rain) do play an important role, a profound knowledge in cognitive science is needed [17]. The task for the reconstructionist is to determine how much time passed between the first possible moment for the driver to realize the hazard and the beginning of the evasive maneuver.

In [7], the triplet "perception–decision–reaction/response" is abbreviated by the acronym PDR, and the time needed to perform these tasks is termed PDR time. PDR times may vary widely and depend on numerous factors including weather and traffic environment, day- or nighttime, type of vehicle, and mental and physical condition of the driver (alcohol, drugs, medication). The perceptibility of what constitutes a hazard for the driver is of utmost importance. A car entering an intersection on a sunny day is much easier to perceive compared to a pedestrian attempting to cross a road coming from behind a parking car during nighttime. Consequently, the PDR time can range from 0.5 to about 2.5 s. In [11], the authors conclude that, under normal circumstances, it takes a car driver on average between 0.8 and 1.0 s to set an avoidance action. This is the time period from the instant when the phase of perception is finished and the driver is aware of a hazard and the necessity of a reaction, until the brakes or any other avoidance action becomes effective.

During PDR time, the driver also has to make a decision about the action for accident avoidance. Under normal circumstances, a car or truck driver will have no choice other than a braking maneuver. A motorcycle rider, however, may also try an evasive lane-change maneuver to circumnavigate an obstacle blocking his or her way. We will address motorcycle-related issues in a later chapter and assume that the driver initiates an emergency braking maneuver for accident avoidance.

41.3.1.1 Calculation of a Basic Braking Maneuver

It is commonly assumed that within PDR time the dynamic properties of a vehicle play a minor role. Consequently, the vehicle is modeled as a single mass with one degree of freedom (DOF) in

longitudinal direction. This simplification allows to use the formulas listed in the following for all kinds of vehicles. Nevertheless, it is necessary to make sure that this is a valid model for the pre-crash phase. Since there is almost always a lack of information concerning the acceleration during that phase, frequently, a constant value is assumed, making it is easy to calculate unknown parameters for a given set of known or assumed other parameters. In Equations 41.1 through 41.5, as known from textbooks as well as from [8], the following symbols are in use: t for time, d for distance, v_i and v_e for initial and for end velocity, respectively, and a for a constant acceleration. A negative value has to be used for a to become a deceleration.

To calculate the time t to accelerate/decelerate from an initial speed v_i to a final speed v_e at a constant acceleration a,

$$t = \frac{v_e - v_i}{a}. \tag{41.1}$$

To calculate the travelled distance d when accelerating/decelerating from an initial speed v_i to a final speed v_e or within a time period t at a constant acceleration a,

$$d = \frac{v_e^2 - v_i^2}{2a} = v_i t + a\frac{t^2}{2} = \frac{t}{2}(v_i + v_e). \tag{41.2}$$

To calculate the final velocity v_e from an initial speed v_i for a given time t, distance d, or constant acceleration/deceleration a,

$$v_e = v_i + at = \sqrt{v_i^2 + 2ad}. \tag{41.3}$$

To calculate the initial velocity v_i for a constant acceleration/deceleration a from a final speed v_e and/or a given time t and distance d,

$$v_i = v_e - at = \frac{2d - at^2}{2t} = \sqrt{v_e^2 - 2ad}. \tag{41.4}$$

To calculate the constant acceleration a from an initial speed v_i and for a given time t, distance d, or a final speed v_e,

$$a = \frac{v_e - v_i}{t} = \frac{2d - 2v_i t}{t^2} = \frac{v_e^2 - v_i^2}{2d}. \tag{41.5}$$

If the accident can be avoided, in general the final velocity v_e would be zero. Otherwise, v_e is the collision speed and will be needed in the subsequent crash analysis. If the acceleration is not constant for the entire maneuver, suitable formulas need to be derived, or the maneuver may be subdivided into sections of almost constant acceleration, for which the presented equations can be applied. The entire braking maneuver is then composed by assembling the individual results.

41.3.1.2 Estimation of Braking Deceleration

The choice of value used for the acceleration a depends on several factors, such as type of vehicle, braking system, degree of brake application, type of road, and road condition. Only for a full braking maneuver with all wheels locked or operating near the optimum slip (see Chapters 17–19), the deceleration can be related directly to the friction parameter of the combination of road and tire. Nowadays, practically all new passenger cars are equipped with an antilock braking system (ABS), which guarantees almost optimal deceleration, if engaged by the driver for the entire braking maneuver. The prevalence of ABS is not nearly as high, however, for other vehicles like trucks

and powered two-wheelers (PTWs) and makes estimating the deceleration more difficult. In principle, there are three approaches available to put such estimated values on a solid base:

1. Using measured longitudinal forces and friction parameters for tires on the kind of surface and for conditions equivalent to that at the accident scene
2. Using measured deceleration values from vehicle manufacturers, consumer reports, or specific references
3. Making individual measurements at the accident scene, preferably with a vehicle of the same make and year under similar road conditions

Approach (1) has the advantage of not being restricted to a specific vehicle or braking equipment. Based on more elaborate vehicle models, as presented in other parts of this book, for example, the brake force distribution between axles can be taken into account, and quite realistic results can be obtained. Although there are numerous references of measurements in the tire literature, they are mostly valid for artificial surfaces, as used in tire testing machines. One will hardly find data for realistic surfaces, like a used and wet asphalt pavement or for icy road conditions.

The second approach (2) is the most common method to estimate the achievable braking deceleration. Measured data from vehicle tests can be found in various comprehensive sources like [8,10,11,18] but also in journal articles and vehicle reports. Note that parameters of the testing method and the method of data analysis have an influence on the results. Therefore, it is necessary to closely examine such data in order to avoid a misinterpretation. For instance, in [8], the range for nondimensional deceleration under various conditions is given for different initial speeds (below and above 30 mph) of the braking maneuver. On the other hand, in [11], it is pointed out that 50% of the brake response time is included in the results. Table 41.1 is a compilation from several references listing maximum braking deceleration achieved with passenger cars on various road surfaces and road conditions. Differences within testing methods and parameters are accounted for by listing ranges for the results. Note also that results from different sources vary widely, especially for low friction surfaces, and original references need to be consulted when data are used at court.

Occasionally, approach (3) is chosen, especially in difficult cases. Although this seems to be the most reliable method, there are still uncertainties that are hard to overcome. Road and tire temperature and tire age and road condition have a significant influence on the generation of braking forces. Since these parameters are rarely known, it is virtually impossible to reproduce exactly a braking

TABLE 41.1
Maximum Nondimensional Braking Deceleration[a] Achieved with Passenger Cars

Road Surface and Condition	Dry	Wet
Concrete new	0.60–1.00	0.50–0.75
Concrete used	0.50–0.75	0.45–0.65
Asphalt new	0.60–1.00	0.50–0.75
Asphalt used	0.50–0.70	0.45–0.65
Gravel packed	0.35–0.70	0.40–0.70
Gravel loose	0.30–0.70	0.45–0.60
Grassy ground	0.30–0.60	0.10–0.40
Snow packed	0.30–0.55	0.30–0.60
Snow loose	0.10–0.25	0.05–0.25
Ice	0.10–0.20	0.02–0.10

[a] a/g with g being the acceleration due to gravity.

maneuver that has happened in the past. However, chances are good to be as close to reality as possible, and valuable insight into the accident may be gained from on-site testing. On the other hand, application of this approach is also limited because it might be impossible to carry out such tests for various reasons.

41.3.1.3 Analysis of Evidence at the Accident Scene

Aside from the uncertainty of the technically achievable braking deceleration, there is also the question of brake application by the driver. Unless skid marks are found, there is no evidence how heavy, or if at all, the brakes have been operated. Skid marks are an important piece of evidence, and therefore reconstruction of the precrash phase relies heavily on them. With recent cars and ABS, however, skid marks become rare, since wheel locking is prohibited by the braking system. Although a braking wheel at optimum longitudinal slip may leave some traces on the road surface, such tire marks, if present at all, are difficult to recognize and easily overlooked when the scene of the accident is examined. The resulting lack of evidence of a braking maneuver with ABS-equipped vehicles is an additional difficulty in an accident analysis. Therefore, event data recording devices (EDS) that record and store vehicle data like speed and brake application prior to an accident will be the only reliable solution to this problem for the future [19].

In the discussion of the precrash phase, it was assumed that the driver has the vehicle under control until the crash occurs. The only input assumed for the driver was the application of the brakes. There are, however, accidents where the driver loses control over his or her vehicle before it collides with other vehicles or objects. Typical cases are rollover accidents and accidents where the vehicle unintentionally goes off the road. Evidence is mostly found as curved and diverging tire marks resulting from tires operating at large slip angles. Such accidents need to be analyzed differently already in the precrash phase, and a more elaborate model must be employed to capture the precrash dynamics of the vehicle.

41.3.2 Crash Phase

Compared to the pre- and the postcrash phase, the crash phase itself is very short lived. Nevertheless, it is the phase that makes the incident an accident, with consequences for the concerned persons and involved real assets, ranging from minor to dramatic. The crash phase is characterized by the action and exchange of mostly very high contact forces within a short period of time. It starts with the first contact of the colliding bodies and ends when the contact is terminated. A typical collision of a car with another car or a road barrier lasts for only about 100–200 ms and is also called *impact*. During this short interval of time, the contact forces cause deformations of the impacting bodies. These deformations can be nonpermanent or permanent, that is, *elastic* and *plastic*, respectively. While for an elastic impact of two bodies, both momentum and kinetic energies are conserved, in the case of a nonelastic (plastic) impact, some kinetic energy is transformed into deformation.

There are several excellent literary sources that treat impact dynamics of bodies comprehensively and in a general manner (see, e.g., [20,21]). Sources that are focused on the impact and collision of vehicles are the books [7,22,23]. But practically all sources on accident reconstruction comprise parts that deal with the mechanics of impacting and colliding vehicles.

To outline basic impact mechanics of colliding vehicles, one of the most simple accidents, a straight-line rear end collision (see Figure 41.1), will be discussed now. We assume that two vehicles of masses m_1 and m_2 are moving at velocities v_1 and v_2 in the same direction, with vehicle m_2 being ahead and $v_1 > v_2$. Centers of gravity of the vehicle masses and velocity vectors are aligned, and so a 1D impact can be considered. At time $t = t_b$, contact and compression and deformation of the vehicle structures start. This first phase, called compression phase, ends at $t = t_m$ when the minimum distance between the centers of gravity is reached. Both vehicles now move with identical velocity v_m.

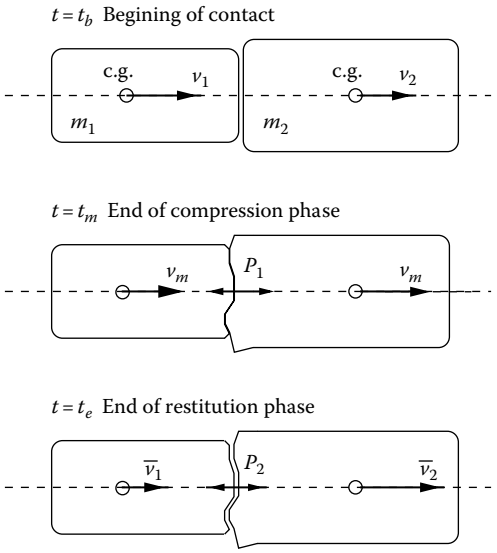


FIGURE 41.1 Definition of phases of a 1D impact.

From the conservation of momentum, it follows that at the end of the compression phase, an exchange of impulse P_1 has taken place between vehicle m_1 and m_2:

$$m_1(v_m - v_1) = -P_1 \quad \text{and} \quad m_2(v_m - v_2) = P_1. \tag{41.6}$$

A second phase, the restitution phase, follows, where deformation back-forming to some degree may take place and the vehicles move apart until the contact ends at $t = t_e$. At t_e, vehicle m_1 moves with velocity $\overline{v}_1$, and vehicle m_2 with $\overline{v}_2$ and $\overline{v}_2 \geq \overline{v}_1$ holds. Again, an exchange of an unknown impulse P_2 has taken place:

$$m_1(\overline{v}_1 - v_m) = -P_2 \quad \text{and} \quad m_2(\overline{v}_2 - v_m) = P_2. \tag{41.7}$$

Merely four equations (41.6 and 41.7) are available for the five unknown quantities v_m, $\overline{v}_1$, $\overline{v}_2$, P_1, and P_2. This dilemma can only be solved by an additional equation, that is, obtained from Newton's hypothesis, suggesting that a restitution coefficient k exists that relates P_1 to P_2:

$$P_2 = kP_1 \quad \text{or} \quad (\overline{v}_2 - \overline{v}_1) = k(v_1 - v_2). \tag{41.8}$$

Now, the final velocities $\overline{v}_1$, $\overline{v}_2$ after the impact can be calculated from

$$\overline{v}_1 = v_1 - \frac{(v_1 - v_2)(1+k)}{1 + m_1/m_2} \quad \text{and} \quad \overline{v}_2 = v_2 - \frac{(v_2 - v_1)(1+k)}{1 + m_2/m_1}. \tag{41.9}$$

Introducing a restitution coefficient k is a simple trick to overcome the mathematical problem of a missing equation but raises another difficulty: what value shall be inserted for k? When $k = 1$, a fully elastic impact is assumed. This may be valid in real accidents for collisions at very low speeds where only the bumpers are elastically deformed and fully restitute. The other extreme $k = 0$ (fully plastic impact) may hold for a high-speed collision with vehicles deformed and crushed beyond recognition. Obviously, k can be related to the relative velocity of the colliding vehicles (see [24]). Another possibility for the reconstructionist is to make a good guess, mainly based on the

deformations of the vehicles. Well-documented vehicle crash tests are probably the only way to gain sufficient experience and a solid base to make an educated guess when the coefficient of restitution needs to be determined. Some sources that may be helpful are [24] (www.crashtest-service.com) for the Old World and [18] for the New World. This differentiation is indicated, since vehicle models are still different in Europe and in the Americas.

While the total momentum is unchanged for the system of two colliding bodies and does not depend on the restitution coefficient, there is an energy loss for $k < 1$. This energy loss E_C of a partly inelastic impact can be calculated from

$$E_C = \frac{1}{2}[(m_1 v_1^2) + (m_2 v_2^2)] - \frac{1}{2}[(m_1 \overline{v}_1^2) + (m_2 \overline{v}_2^2)], \tag{41.10}$$

and with the solutions Equation 41.9, it follows

$$E_C = \frac{1}{2}(1-k^2)\frac{m_1 m_2}{m_1 + m_2}(v_1 - v_2)^2. \tag{41.11}$$

As one can see, this equation relates the dissipated energy to the coefficient of restitution. Therefore, one can also try to estimate the deformation energy, instead of guessing k, and solve the missing equation problem. In fact, this is another approach that became popular at the time when an almost linear relationship between residual deformation and impact speed [25] was found. Since then, large efforts have been made to analyze crash tests against barriers, obstacles, and other vehicles in order to determine the deformation (crush) energy and to transfer the results to other collisions of the same vehicle with similar deformations (see [26] and many other papers of this kind in [18]). Also, we will come back to this topic in Section 41.4.3.2.

The straight-line rear end collision was also used as the example to discuss the crash phase since it is an accident scenario that is frequently associated with whiplash injuries of the passengers in the vehicle in front. This kind of injuries has become very common in the past, partly because there is practically no visible evidence of this cervical spine trauma, making it very hard to prove or disprove the existence [27]. In particular, the accusation of fraud is associated with this injury. The accident reconstructionist is challenged since the probability of a whiplash injury of the passengers in a vehicle is linked to the velocity change $\Delta v = \overline{v} - v$. Therefore, the analysis of this seemingly rather simple to analyze accident scenario can become very tricky (see [28,29] for further discussion of biomechanical aspects). On the other hand, the automotive industry has also reacted, and, for example, the seat design is continuously improved to reduce the risk of whiplash injuries in rear accidents (see [30,31]).

41.3.3 Postcrash Phase

This phase starts when the last contact between crash opponents has ended and a vehicle or accident victim moves into its final position. Occasionally, it may be difficult to draw a clear line between the crash and the postcrash phases, especially with secondary and minor impacts after the main collision. The postcrash phase ends when the vehicle or its driver comes to rest. It is important to realize that vehicle and driver may separate during the crash phase and therefore their postcrash phases may be independent from each other. This occurs quite frequently for PTW's and their riders, but this may also happen to drivers of convertible cars. In rare cases, even drivers of regular cars are ejected, mostly when the safety belt was not used and the doors opened during the crash.

The final or end positions of the vehicles involved in an accident are often known, especially when vehicles are not mobile after a severe accident. Additionally, skid marks, scratches, and other traces give evidence of the path the vehicle did follow during this phase. Together with other evidences on the location of the crash, like kinks in skid marks, traces of fluids on the ground, pieces of

glass, or debris, it is quite often possible to identify the beginning, the end, and the path in between of a vehicle's postcrash phase. For pedestrians, drivers, or passengers, mostly, only the final position is known, since significant parts of the postcrash trajectory may be travelled in the air or at least without sufficient contact to the ground to create traces.

Since the postcrash phase generally is the one to offer the most reliable information and also the information suitable for an evaluation, the accident reconstruction frequently starts with the analysis of this part of the accident event. By going backward in time starting from the final position with all velocities being zero, the velocities at the beginning of the postcrash phase are sought. By employing a suitable mathematical model, it may be possible to find a numerical solution for the motion of the vehicle. Ideally, such a solution would reproduce the velocities at the beginning and also match with evidence (skid marks, spilled fluids, location of debris etc.) at the accident site.

In general, the motion of a vehicle in the postcrash phase will be a combination of longitudinal, lateral, yaw, and roll motions. Only in simple accident scenarios like a straight-line rear end collision, or a car hitting a pedestrian or a lightweight object, the longitudinal motion will be predominant. But even in these cases, a significant pitch motion may occur. The point is that other than for the precrash phase, the application of very simple models like the single DOF model as used in Section 41.3.1 is quite restricted, or, if applied nevertheless, the value of the results may be limited. Therefore, more elaborate models are needed that capture the postcrash motion in more details. For accidents with vehicles that do not roll or tip over, a model capable of describing the planar motion by the longitudinal, the lateral, and the yaw motion may be sufficient. If the roll motion is significant or if overturning occurs, then a 6-DOF model will be needed.

Two-wheeled vehicles are especially difficult to model during this phase, since overturning can be an essential part of the postcrash phase. It may occur simultaneously within the crash, immediately thereafter or after travelling some distance in a stable upright position.

Likewise, the postcrash motion of human beings and animals is quite complex. Phases of free flight with full translational and rotational motion of the main body, plus additional motion of extremities, are disrupted by contact with the ground or objects. In addition, possible muscular activity during this phase may affect the postcrash trajectory significantly. Due to the complexity of valid models and also the unknown inputs, in general, the analysis mainly relies on experimental data supplemented by an analysis based on very simple models [32]. There is, however, multibody system software for biomechanical systems available to simulate the complex motion of pedestrians and vehicle occupants to some extent (see Chapter 5). The application of such software may be indicated in special cases of an accident reconstruction.

41.3.4 Reconstruction Strategies

A common and main goal of an accident reconstruction is to determine trajectories and velocities of all participants and involved objects in the accident. This will enable the reconstructionist to answer questions about the speed and path driven by the vehicles involved and may explain the cause of a traffic accident and the liability of the participants.

The strategy to solve the reconstruction problem largely depends on the information and evidence available. First, one will try to determine vehicle positions and velocities at the beginning and at the end of the phase for which most data are accessible. Thereby unknown parameters will become available for an adjacent phase, and so the next phase can be dealt with. Since this approach separately treats each phase, it is possible to use different calculation methods and also different models for each phase. Ultimately, all three phases are solved for all participants in the accident and can be stitched together for a unified solution. This approach has the advantage that methods and models to analyze a phase can be selected according to the specific characteristics of an accident.

A different approach is used with some software tools (see Chapter 5) where numerical simulation is employed to solve the governing differential equations of the vehicle model. Due to the nature of this approach, initial state values and impact parameters have to be determined or assumed, and

then the sequence of phases is computed from beginning to end. The crash phase as the major discontinuity in this process is handled based on data provided initially or interactively. When the end position of a vehicle is reached, computations will be stopped. Now, the results have to be compared with information not used so far, for example, the final position of a vehicle. If the known (documented) final position matches the computational results reasonably well, and if other evidences fit in, then it can be assumed that a valid solution was found. If there are discrepancies, the assumptions need to be changed, and another try can be started as a second step in an iterative procedure that should converge toward the solution satisfying all verified information and evidence.

41.4 MODELS FOR ACCIDENT ANALYSIS

Mathematical vehicle models for the analysis and reconstruction of vehicle accidents are, in principle, not different from the models used in other investigations of vehicle dynamics. There is, however, the need for rather small models for several reasons. First of all, with increasing complexity, also the number of necessary input parameters increases. Often, such data are not available and cannot be investigated. Consequently, assumptions need to be made, which can make a more accurate model less faithful and diminish the original advantage of the detailed model. Secondly, such models must not cause undue computational time on a desktop computer. This aspect is important since calculations need to be carried out numerous times if the solution can be found only by an iterative procedure.

Another issue with accident analysis models is the area of validity with regard to the range of certain parameters. Since the motion of a vehicle can be quite extreme, especially in the postcrash phase, fully linear models cannot be used in most cases. Dynamic parameters of the vehicle like tire slip angles may take on any value between 0 and $\pi/2$ during skidding, and a linear tire characteristic would be inappropriate.

In the following sections, an overview will be given on the modeling of vehicles, particularly with regard to passenger cars. The content focuses on the main modeling issues, the vehicle as a whole, the tire characteristics, and the collision event. Considering other vehicles like motorcycles and semitrailers would go beyond the scope of this chapter, and one may check the listed literature for details on other vehicles.

41.4.1 Vehicle Models

Typically, vehicle models with a low number of DOFs are in use. Either the analytical solution of the differential equations describing the model is employed or the solution of the equations is sought by a numerical time-stepping routine.

41.4.1.1 Kinematic Models

Kinematic models are based on 1-DOF models of a single mass. If the acceleration of mass m is assumed to be a constant parameter a, the governing equation of a linear motion in direction x becomes

$$m\ddot{x} = ma = \text{const.} \tag{41.12}$$

and can be solved analytically on the time interval $t = [t_i, t_e]$. For a given initial distance x_i and initial velocity v_i at initial time $t_i = 0$, one obtains

$$\dot{x}(t_e) = v_e = at_e + v_i \quad \text{and} \quad x(t_e) = d = a\frac{t_e^2}{2} + v_i t_e + x_i. \tag{41.13}$$

These basic equations can be converted, and additional relations can be derived as shown in Equations 41.1 through 41.5. Many fundamental problems in accident analysis can be investigated by applying these equations. This approach can also be used for the motion on a circular path,

lane-change maneuvers, and others. Note that the key point and essential simplification in this example is to assume the acceleration a to be constant within the time interval $t = [t_i, t_e]$. Different assumptions, for example, a linearly increasing acceleration or a constant velocity, can be prescribed as well, and analytical solutions derived.

It is important to realize that the forces needed to bring about a certain enforced motion do not appear in the solution of Equations 41.1 through 41.5 and 41.13. Therefore, one has to verify that implied conditions on, for example, braking or cornering forces do not exceed physical limits.

The advantage of such kinematic models is their simplicity, which makes coding easy and facilitates the usage in spreadsheet programs and other software of that kind. The drawback clearly is that only simple cases of motional behavior can be analyzed and kinetic models must be used for more complex situations.

41.4.1.2 Kinetic Models

Kinetic models, by some authors also termed dynamic models, are represented by differential equations for single- or multibody systems based on Newton's laws. These equations of motion are well known from other chapters and are repeated for completeness in a general matrix notation:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} = \mathbf{F}(\dot{\mathbf{q}}, \mathbf{q}, t). \tag{41.14}$$

Vector $\ddot{\mathbf{q}}$ is the second time derivative of the generalized position vector $\mathbf{q}$ of length n, which contains translational and rotational displacements, associated with the n DOFs assumed for the system. Typically, 3-DOF planar systems with $\mathbf{q} = [x,y,\phi]^T$ are used for vehicles models, allowing for longitudinal and transversal motion $\dot{x}$, $\dot{y}$ and the yaw motion $\dot{\phi}$ (see [7,22]). Very few 6-DOF models that account for the full motion of the main vehicle body are employed. Complex multibody models for single vehicles, as frequently applied in vehicle dynamics analysis, are rarely used, due to reasons as outlined before. Vehicle combinations, for example, trailers attached to trucks or cars and articulated vehicles, need, of course, models with additional DOFs.

In Equation 41.14, matrix $\mathbf{M}(\mathbf{q})$ represents the mass matrix of the system, containing the mass and the rotational inertia properties of the system. For a rigid-body vehicle, this matrix is constant $\mathbf{M}(\mathbf{q}) = \mathbf{M}$; however, with multibody models, it may depend on the system states.

The complexity of such models, even for 3-DOF planar systems, lies in the right-hand side of Equation 41.14, which contains all external forces acting on the system and gyroscopic forces as well, if present. Noticeable external forces during the pre- and the postcrash phase result from aerodynamic forces $\mathbf{F}_{aero}(\dot{\mathbf{q}}, \mathbf{q})$ and from tire forces $\mathbf{F}_{tire}(\dot{\mathbf{q}}, \mathbf{q}, t)$. The latter constitute the most dominant forces that determine almost exclusively the motion of the vehicle during the noncrash phases. Tire forces are determined from mathematical models describing the tire behavior under various conditions and will be discussed in the next section.

Aerodynamic forces depend largely on the speed of the vehicle since they increase with the square of the relative velocity to the air mass. For high-speed accidents, they may be important and can be taken into account by respective forces and moments acting along the principal axes of the vehicle body. These forces and moments can be calculated from aerodynamic coefficients like for drag and lift (see [33]).

41.4.2 Tire Force Models

Except during the crash phase, tire forces are the most important forces to operate on the vehicle during a vehicle accident. Therefore, its motion is essentially determined by these forces, and their accurate calculation is of utmost importance. It is the role of a tire force model to calculate the forces and moments acting within the tire–road contact area and provide them for the vehicle model. In the past, numerous tire force models have been developed and reported. A recent compilation of such models is found in [34] and a model performance test in [35]. A suitable tire force model, however,

must be tailored to the particular needs of accident reconstruction. The main requirements are the validity for large values of longitudinal slip and of side slip, scalability to suit different tire sizes, ease of application, and, last but not least, fast numerical calculation.

Generally speaking, the tire force model is a mathematical function that computes the forces and torques needed for the vehicle model from a set of input parameters:

$$\mathbf{F}_{tire} = [F_x, F_y, M_y]^T = \mathbf{f}(F_z, \dot{x}_T, \dot{y}_T, \dot{\phi}_T, \omega_T, p_1(t), t, \ldots, cp_n(t)). \tag{41.15}$$

Equation 41.15 is a generic formulation of such a model, which computes the longitudinal and lateral tire forces F_x and F_y, respectively, as the most important forces needed. To account for the rolling resistance, the moment M_y may also be computed. The other moments M_x and M_z acting on the tire are rarely applied to accident reconstruction models, except for single-track vehicles.

Note that the vertical tire force F_z is an input parameter to the model. Other input parameters are the angular velocity ω_T and the longitudinal and lateral velocities $\dot{x}_T$, $\dot{y}_T$ that determine the actual longitudinal and lateral slip as the characteristic operating parameters of a tire. Some input parameters $p(t)$, like a steering input by the driver, may depend on time t; others may be assumed constant in certain regions of the solution domain, for example, the tire–road friction parameters; and some are simply numerical parameters of the model. Due to the nonlinear behavior of the tire, especially when operated at limit situations, most tire force models cannot be represented by just one universal function **f** as it may appear from Equation 41.15. In fact, the tire force model **f** may consist of a set of s functions $\mathbf{f}_{1,\ldots,s}$ that account for different ranges of the operating parameters.

In regard to the complexity of most tire force models on one hand and the comprehensive discussion of the tire as a vehicle component in Part III of this handbook on the other, a brief overview on current state-of-the-art tire force models is given in the following. It is based on a very detailed treatise [36] on various tire force models as currently used in vehicle accident reconstruction software. Since elaborate tire force models are needed primarily in accident simulation software, there is a strong tie between such software and tire force models.

41.4.2.1 "Magic Formula" Tire Force Model

This tire force model is comprehensively discussed in Chapter 17 and in [37]. In the literature originating from the New World, the model is also referred to as the Bakker–Nyborg–Pacejka (BNP) model, because of Ref. [38]. The *magic formula* described therein is capable of calculating the characteristics of side force, brake force, and self-aligning torque. The mathematical representation is limited to steady-state conditions during either pure cornering or pure braking and forms the basis for a model describing tire behavior during combined braking and cornering. Its versatility can be seen from the fact that it can be not only used for general vehicle dynamic analysis but also adapted and used for vehicle accident reconstruction.

41.4.2.2 VCRware Tire Force Model

This tire force model is based on a subset of the BNP model [38]. For combined longitudinal and lateral forces, the Nicolas–Comstock–Brach (NCB) equations are employed (see [7,39]). A characteristic feature of the model is that it consists of functions for the longitudinal and the lateral tire forces $F_x(\alpha, s)$ and $F_y(\alpha, s)$ that are continuous functions over the entire range of the lateral slip angle ($0 \le \alpha \le \pi/2$) and the longitudinal slip ($0 \le s \le 1$) from a free rolling to a locked wheel. This tire force model is used in the accident simulation software VCRware (vehicle crash reconstruction software) (see Section 41.5).

41.4.2.3 PC Crash Linear Tire Force Model

The accident simulation software PC Crash (see Section 41.5) offers and works with two different tire force models, a "linear" model and the TMEasy tire force model (see the next section). The "linear" model is not an entirely linear model, but it approximates the nonlinear tire characteristics by piecewise linear and almost linear functions and is hence termed "linear model" within PC Crash.

Longitudinal and lateral tire forces are defined within three regions of the lateral slip angle α. The lateral force F_y increases linearly within the first region ($0 \leq \alpha \leq \alpha_{max}$) and remains constant at the friction limit in the second region beyond α_{max}. The longitudinal tire force F_x is at the maximum value in region I and II. Region III begins when F_x and F_y both reach the friction circle. Both forces are linked together in region III by the equations for the friction circle for a fully skidding tire with slip angles $\alpha \leq \pi/2$. See [22] or [36] for a full discussion of the modeling aspects.

With less equations to evaluate and less harmonic functions to compute, this model has the advantage of fast numerical computation. Of course, there is a price to pay. Compared to other models, the approximation of the tire characteristics is of lesser quality in certain regions of the parameter range [36].

41.4.2.4 TMEasy Tire Force Model

The TMEasy tire force model is presented in full details in [40,41]. It is a rather sophisticated semiphysical tire model that is suitable for planar motion as well as for full 3D motion of a vehicle. TMEasy calculates the contact force characteristic in longitudinal and lateral direction from a rather small number of physical parameters. The combined force characteristics are directly generated via a generalized slip approach. With this model, also the tire loads as the third component of the tire contact forces can be computed. In addition, and not available with most other models, self-aligning moment, rolling resistance moment, and also the bore moment (torque) can be determined.

Furthermore, another feature distinguishes the TMEasy model; it includes nonlinear first-order dynamics for longitudinal and lateral forces and for the aligning moment. From the viewpoint of model fidelity, this is a clear advantage of this model, since highly transient vehicle motions regularly occur during postcrash motion.

The integration of TMEasy into other software is facilitated by a standard tire interface (STI). TMEasy and some of its derivatives are used in various simulation software for vehicle dynamics, among them the accident reconstruction program PC Crash. Taking advantage of all features of the tire model only makes sense with an adequate vehicle model and results in comparably slow execution time.

41.4.2.5 Other Tire Force Models

Due to the need of an appropriate tire force model in a vehicle accident simulation program, each of these programs has their own tire model, which is mostly developed as variations of previously existing models. The software SIMON (see Section 41.5.1) uses a semiempirical model based upon the widely used HSRI tire model [42], which is known for quite a long time. Another program, SMAC, uses a tire force model, which has its root in an even older model by Fiala [43] and is discussed in [7].

41.4.2.6 Evaluation and Comparison of Tire Force Models

This list, although incomplete, shows that there are quite a number of different tire force models in use. As pointed out already, the tire forces determine the vehicle motion during the postcrash phase to a great extent. Therefore, tire force models play an important role in the process of analyzing and reconstructing a vehicle accident by numerical methods.

Nevertheless, there is no procedure available to validate such tire models or to check them for correctness. Crash tests have been used in the past to test and benchmark simulation programs, but such comparisons involve the entire vehicle model including the tire model. In [36], a first attempt was made to highlight the influence of tire models. The use of simulation models with different tire models but identical initial conditions led to quite different final positions after postcrash motions under various conditions of locked wheels and brake application. In the interest of comparable and reproducible results, establishing benchmarks for tire models and standards for the definition of tire parameters should be a goal for the future.

41.4.3 Crash Models

During a vehicle-to-vehicle collision, forces that are several magnitudes higher than during the pre- or the postcrash phase act in the contact area. Moreover, this happens typically within a few tenth of a second. Therefore, compared to the other phases, the crash phase needs to be treated differently.

The goal of a crash model is to calculate the states of the vehicle(s) immediately after the crash, based on the states when the crash phase starts. There are two major and quite distinct approaches possible:

- Models based on the mechanics of impacting rigid bodies, which implies an infinitely short contact time with infinitely high forces acting in the contact interface
- Models that resolve the contact in the time domain, calculate the contact forces, and apply them to the colliding objects

The first approach has the clear advantage that only a few algebraic equations need to be solved. Modeling the collision of vehicles by impacting rigid bodies, however, implies stringent conditions, which either limit the applicability considerably or are ignored in an undue manner. The second approach certainly captures the actual mechanical processes much better and gives more realistic results. The drawback is the enormous computational time for really detailed models. In addition, all but the manufacturer of a vehicle will mostly fail with this approach due to the lack of specific data needed, if one wants to take full advantage of the model. However, in recent years, there has been progress made in developing mesh methods that are a compromise between model fidelity and numerical effort (see Section 41.4.3.3).

The topic of modeling the collision of vehicles is covered by extensive literature. Comprehensive references directly related to the reconstruction of vehicle accidents are [7,10,22,23,44] and others. For an in-depth study of the topic, these books will be valuable sources.

41.4.3.1 Planar Impact Mechanics

As a first example of impact mechanics, the 1D impact between vehicles was already discussed in Section 41.3.2. Here, the most widely used governing equations for planar impact mechanics of vehicles are presented, and the basic assumptions are discussed. As outlined, impact models consist of a set of algebraic equations that constitute a relation between the linear and the angular velocities v, ω before the impact and $\overline{v}$, $\overline{\omega}$ afterward. Part of this set of equations are the well-known planar impulse–momentum equations with the change of momentum on the left-hand side and the contact impulse components P_n and P_t on the opposite side:

$$m_1(\overline{v}_{1n} - v_{1n}) = P_n, \quad m_1(\overline{v}_{1t} - v_{1t}) = P_t, \tag{41.16}$$

$$m_2(\overline{v}_{2n} - v_{2n}) = -P_n, \quad m_2(\overline{v}_{2t} - v_{2t}) = -P_t. \tag{41.17}$$

For a general planar motion (see Figure 41.2), where the rotational motion of the colliding bodies is also considered, additional equations can be derived for the change in angular momentum for each vehicle that must be equal to the moment of the contact impulse:

$$I_1(\overline{\omega}_1 - \omega_1) = P_t n_1 - P_n t_1, \tag{41.18}$$

$$I_2(\overline{\omega}_2 - \omega_2) = -P_t n_2 + P_n t_2. \tag{41.19}$$

With six equations for eight unknown quantities $(\overline{v}_{1n}, \overline{v}_{1t}, \overline{v}_{2n}, \overline{v}_{2t}, \overline{\omega}_1, \overline{\omega}_2, P_n, P_t)$ in the general planar case, now two equations are missing (one more compared to the 1D case) to solve the problem.

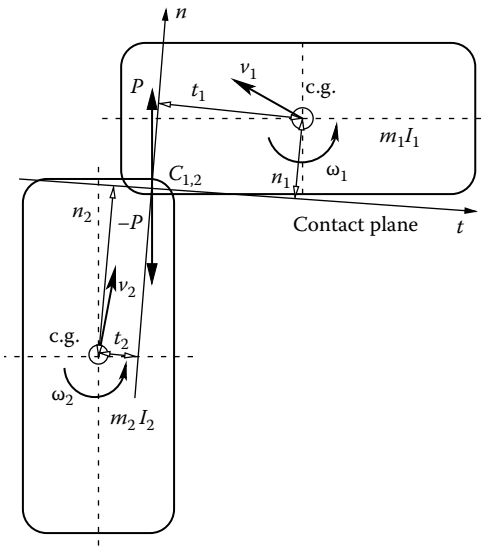


FIGURE 41.2 Planar impact mechanics of two colliding vehicles.

With the contact impulse vector $\mathbf{P} = [P_n, P_t]^T = \mathbf{P_1} + \mathbf{P_2}$, we can again relate the restitution impulse $\mathbf{P_2}$ to the impulse during the compression phase by a coefficient of restitution k and obtain

$$\mathbf{P_2} = k\mathbf{P_1}. \tag{41.20}$$

By eliminating the restitution impulse, we get components in normal and tangential direction to the contact plane for the contact impulse:

$$P_n = P_{1-n}(1+k), \tag{41.21}$$

$$P_t = P_{1-t}(1+k). \tag{41.22}$$

Obviously, there are still too many unknowns, and a second hypothesis is needed. It will be a kinematic assumption concerning the relative velocity $\Delta\mathbf{v_c} = [\Delta v_{c-n}, \Delta v_{c-t}]^T$ of the congruent contact points C_1,C_2 on the vehicles 1 and 2. This relative velocity is

$$\Delta v_{c-n} = v_{1n} - \omega_1 t_1 - (v_{2n} + \omega_2 t_2), \tag{41.23}$$

$$\Delta v_{c-t} = v_{1t} + \omega_1 n_1 - (v_{2t} + \omega_2 n_2). \tag{41.24}$$

We will assume that the direction of this relative velocity at the beginning of contact will be constant within the compression phase:

$$\frac{\Delta v_{c-t}}{\Delta v_{c-n}} = \text{const.} \tag{41.25}$$

This is the essential assumption of the frequently so-called *Kudlich–Slibar impact model* [45]. It is important to realize that the direction of $\Delta\mathbf{v_c}$ is not the direction of the impulse $\mathbf{P}$. Nevertheless, based on Equation 41.25, one can derive the appropriate contact impulse components (see [45]):

$$\Delta v_{c-n} = hP_{1-t} - c_n P_{1-n}, \tag{41.26}$$

$$-\Delta v_{c-t} = c_t P_{1-t} - h P_{1-n}, \tag{41.27}$$

with constants

$$\begin{aligned} c_n &= \frac{1}{m_1} + \frac{1}{m_2} + \frac{n_1^2}{I_1} + \frac{n_2^2}{I_2} \\ c_t &= \frac{1}{m_1} + \frac{1}{m_2} + \frac{t_1^2}{I_1} + \frac{t_2^2}{I_2} \\ h &= \frac{t_1 n_1}{I_1} + \frac{t_2 n_2}{I_2}. \end{aligned} \tag{41.28}$$

By rearranging the equations and inserting Equation 41.21, we get the final result:

$$P_n = (1+k)\frac{\Delta v_{c-n} c_n + \Delta v_{c-t} h}{h^2 - c_n c_t}, \tag{41.29}$$

$$P_t = (1+k)\frac{\Delta v_{c-n} h + \Delta v_{c-t} c_t}{h^2 - c_n c_t}. \tag{41.30}$$

With Equations 41.16 through 41.19, 41.23, 41.24, 41.29, and 41.30, the problem is solvable and the unknown quantities can be calculated.

Most impact models for vehicle collisions are either identical to this model or very similar (see, e.g., [7,22,44]), although the explanation of the underlying ideas may read somewhat different.

As with every model, it is important to be aware of its limitations. Note the resultant impulse ratio:

$$\frac{P_t}{P_n} = \frac{\Delta v_{c-n} h + \Delta v_{c-t} c_t}{\Delta v_{c-n} c_n + \Delta v_{c-t} h} = f \quad \text{hence } P_t = f P_n, \tag{41.31}$$

which relates the tangential impulse to the normal impulse. The assumption Equation 41.25 implies that this ratio can be achieved in the contact area. However, this is only possible if the friction condition is such that the local friction coefficient μ exceeds the impulse ratio $f \leq \mu$. Only then, tangential forces of the necessary amount can be developed in the contact area, and Equation 41.25 is a valid assumption. If this is not the case, a sliding contact may have occurred, and the tangential forces (and hence impulses) are limited by the friction coefficient. Strictly speaking, Equation 41.25 is violated and other assumptions must be made. To get around this difficulty, various strategies have been proposed (see again the literature as mentioned so far).

Another crucial question is the location of the contact point. Ideally, the impact model holds for very small deformations. In practice, the deformations are quite often rather large, and therefore it may be difficult to define the contact "point." Its location, however, defines the distances and arms n_1, n_2, t_1, and t_2 and has a direct and decisive influence on the calculation and the results. This becomes especially important in head-on accidents with a small offset of the colliding vehicles. Such accident scenarios are extremely sensitive to the mentioned geometric parameters and need careful analysis. The model is also not suited well for very large offsets either, because then a sliding collision will occur with no distinct restitution phase at all.

Despite all drawbacks and pitfalls, this crash model or one of its varieties is very popular and used in many software packages for accident reconstruction (see Section 41.5).

41.4.3.2 Energy-Based Methods

The crash model presented in the previous section only accounts for the deformation of the vehicles by the coefficient of restitution. All the information about the permanent deformation of the vehicle and the energy dissipation is condensed in this parameter. It suggests itself to analyze these deformations thoroughly and extract additional information about the collision. Car manufacturers do this regularly during the design process of a new vehicle model, when crash tests are carried out and at the same time large finite element (FE) models of the vehicle are used to simulate the crash test on the computer and bring both results in agreement. For an accident reconstruction, it is almost never an option to do a full scale FE analysis of the crash, since such FE models are not available to the reconstructionist in general and creating them would be very difficult, too time-consuming, and expensive.

In the past, there have been attempts, however, to drastically reduce the complexity of the problem and relate the shape of the deformed vehicle to the energy consumed to produce the residual deformations. Starting out from the fixed rigid-barrier tests carried out with passenger cars at various speeds, the test results suggested an almost linear relationship between residual deformation and impact speed [46]. The studies finally led to a well-known relation [47]:

$$\sqrt{\frac{2E_C}{L}} = c_0 + c_1 D, \tag{41.32}$$

which relates the deformation (crush) energy E_C per unit length L to the residual deformation dD measured perpendicular to the original and undeformed surface of the vehicle body. Two constants c_0 and c_1 are to be determined experimentally, however, and are different for each vehicle and even different for each location (front, rear, and side) of the deformation.

The adaption of these findings to real-world car to car accidents and with approach directions other than head-on led to a method named CRASH3 [48]. The basic idea of the method is to start out from a contour line (horizontal projection) of the vehicle before and after the crash (see Figure 41.3). Typically, the deformations at six points along the destroyed contour are taken, and from these measurements, the crush energy is computed (see, e.g., [47]), from

$$E_C = c_0^2 K_1 + c_0 c_1 K_2 + c_1^2 K_3, \tag{41.33}$$

where

$$K_1 = \frac{L}{2} \tag{41.34}$$

$$K_2 = \frac{L(D_1 + 2(D_2 + D_3 + D_4 + D_5) + D_6)}{10} \tag{41.35}$$

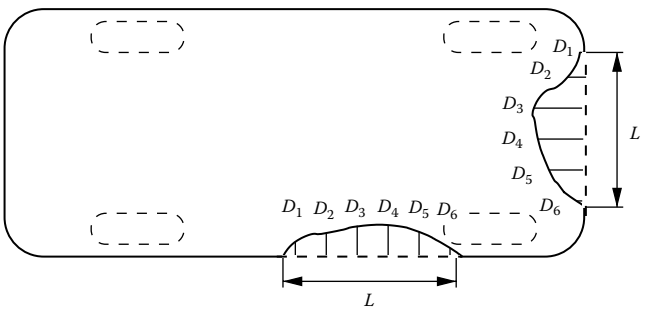


FIGURE 41.3 Estimation of deformation energy from vehicle deformations.

$$K_3 = \frac{L(D_1^2 + 2(D_2^2 + D_3^2 + D_4^2 + D_5^2) + D_6^2)}{30} + \frac{L(D_1D_2 + D_2D_3 + D_3D_4 + D_4D_5 + D_5D_6)}{30}. \quad (41.36)$$

Note that still two parameters c_0 and c_1 need to be known or estimated for a specific vehicle to make use of this method.

Once the deformation energy E_C is calculated, it depends on the kind of impact configuration how to proceed. For a 1D collision with both vehicles moving in the same or opposite direction, the impact problem can be solved immediately (see Section 41.3.2). The deformation energy substitutes for the coefficient of restitution, as introduced in Equations 41.8 and 41.10, can be used as the additional equation needed to complete the momentum equations and to get a solvable system of algebraic equations.

In the case of a general planar impact (2D) or a 3D-impact configuration, the information about the deformation energy is not sufficient to solve the respective impact problem. The equations for the conservation of the linear and the angular momentum need to be supplemented by additional assumptions as this was the case in the previous section. In crash models, frequently, such assumptions are part of the algorithm and not always obvious to the user. A typical example and a general weakness of this approach is the assumption that the entire dissipated energy is equal to the deformation energy. This is especially not true in cases with large frictional forces in the contact area during the collision, like in head-on collisions with a large offset and substantial sliding contact between the vehicles. The CRASH3 model also assumes a perfectly inelastic collision with no restitution of the vehicles perpendicular to the deformed surface. This assumption is also known as the so-called common-velocity condition, since the velocities will be identical at the contact zone after contact for a fully inelastic impact.

Therefore, the velocity changes of the vehicles can be computed in different ways, depending on the details of the crash model (see Ref. [7] for a discussion of this issue). Several such models that calculate the deformation energy are based on the CRASH3 model, which was the first fully developed model of this kind. In the past, modifications and extensions had to be made to make the model more versatile, especially for 3D applications. This aspect will again appear in Section 41.5.

The rather simple method of calculating the deformation energy from measurements at six points along the deformation zone has, of course, called for more sophisticated methods. To mention one of them, in [49], the authors divide the projected deformation area into cells, which they associate with a certain amount of energy necessary to crush this cell. The cell energy data need to be known in advance from crash tests and can be used to estimate the total energy dissipated during the collision for this vehicle. Improved results for the calculated deformation energy can be only expected if the vehicle crash data are adequate.

From these two examples, it already becomes quite clear that evaluating the deformation of vehicles with the goal to calculate the dissipated energy is not a trivial task either. For the most simple model, two vehicle-specific parameters are needed, and more sophisticated models need a lot more data, in addition to the measurements of the deformation area. Therefore, it largely depends on the information available which approach suits best for a certain problem.

41.4.3.3 Alternative Methods

In the two previous sections, two principle approaches of modeling the collision of vehicles were discussed. Both approaches have in common that the collision is treated as an impact, which implies several assumptions. Among them is the presumption that the contact duration is negligible and can be assumed to be zero, which has the consequence that the contact forces become infinite and cannot be calculated directly. For certain accident configurations, this presumption and others involved with the impact model do not hold completely. For example, head-on collisions with a large offset lead to prolonged contact times and significant frictional sliding contact between the colliding vehicles and are basically unsuitable for the impact model.

To overcome the typical deficiencies of the impact model, the only loophole is to resolve the collision in the time domain. As a consequence, rigid bodies cannot be used anymore, and rather than that deformable bodies with appropriate material properties must be introduced. Commonly, this is achieved by some kind of FE approach where the vehicle body or at least the contact area is discretized by deformable structural elements. Another necessity of such models are contact elements on both contacting surfaces to generate the respective contact forces.

The advantage of a more appropriate and reliable model is bought with much more time-consuming numerical calculations. Not only the computational speed is an issue, but also data acquisition and FE-mesh generation add to the additional time needed for an accident analysis. This has led to the development of so-called mesh methods that shall avoid the disadvantages of full FE models but shall be more versatile than simple spring–damper models. An example for this approach is the "DyMESH method" (see [50,51]). It is based on a collision algorithm that uses a triangular mesh defined by the discretized exterior of a vehicle. A possible contact between two meshes is detected by monitoring the positions of nodes of one mesh with respect to the nodes and surface elements of the other mesh. On contact, the interacting 3D forces are generated from vehicle material properties.

Due to the complexity of such mesh models, they can only be used in combination with an elaborate vehicle simulation model (see Sections 41.5.1 and 41.5.2).

41.5 SOFTWARE TOOLS FOR ACCIDENT ANALYSIS

Graphical methods have been in use for a long time in accident reconstruction. But since powerful computers have become available for the reconstructionist, more and more work during an actual accident analysis and reconstruction is performed with the help of computers and software.

The process of an accident analysis and reconstruction involves a number of tasks, starting from the evaluation of evidence including extraction of information from photographs, inspection of vehicles, and data acquisition to the point of the reconstruction of the motion of the vehicles and, where applicable, the accident victims. During almost all of the necessary steps, computers can be helpful in some ways. Therefore, quite a large number of software products are available to the reconstructionist for various tasks [8], covering vehicle dynamics and crash analysis in the first place, but also other important tasks like photogrammetry, image restitution, and calculation of repair costs, to mention a few.

In the following, only software to analyze the motion of vehicles or the collision will be briefly introduced. The primary purpose is to direct the interested reader to some of the most popular software tools. But one might also benefit from this compilation in that it shows the variety of aspects relevant in accident reconstruction. This list of software, however, is incomplete and does not relieve the reader from thoroughly checking what software is available for his or her specific needs. Also, the sequence of the software listed has no special relevance and is simply in alphabetic order.

41.5.1 HVE AND HVE-2D

HVE (human–vehicle–environment) and its constricted version HVE-2D are both products of Engineering Dynamics Corporation (EDC), located in Beaverton, Oregon. The following overview and product description reflect the product's development status when this book was published and is largely based on information and material available from the company's website [52].

Rather than a single simulation program, HVE is a complete 3D simulation environment for setting up and executing simulation modules to study a variety of safety issues related to passenger and commercial motor vehicles, such as dynamic handling behaviors, effects of road conditions, systems failures or driver reactions, and real-world crashes including underrides and rollovers or involving occupants or pedestrians.

By using various program modules specifically designed to take advantage of the capabilities found in HVE, one can perform studies to evaluate the dynamic handling behavior of vehicles, design and test the effectiveness of vehicle braking systems, study driver response to transient effects caused by tire blowouts or irregular terrain, simulate vehicle-to-barrier and vehicle-to-vehicle crash tests, simulate real-world crashes, and identify potential sources of occupant injury resulting from those same crashes.

A number of the so-called editors within HVE allow to move freely between model setup (human, vehicle, and environment), event management, and results. Additionally, HVE offers an extensive feature set available in several HVE-compatible simulation programs. These features allow for advanced simulation of vehicle control systems, driver reactions, tire or brake system failures, and 3D collision modeling. Each HVE-compatible physics program is an independent program running inside the HVE simulation environment.

41.5.1.1 Program Modules Compatible with HVE

The most important models for accident analysis in the HVE environment are listed in the succeeding text. Note that the program modules EDVSM, EDVSD, and SIMON use 3D features and are therefore not compatible with HVE-2D, which is basically a version of HVE that allows only to use planar vehicle models and is restricted to three DOFs for each vehicle:

- *EDCRASH* is a reconstruction program module for single- or two-vehicle accidents, including collisions. The main purpose is to determine the impact speeds and delta-V velocities based on accident site evidence and vehicle damage measurements. This module is also usable for collisions with immovable barriers, such as bridge abutments and large trees. While EDCRASH has been extended and revalidated for use in HVE to account for 3D terrain, it is essentially a 2D reconstruction program.
- *EDSMAC4* is a simulation program module for vehicle collisions. Based on EDSMAC, EDSMAC4 includes numerous extensions developed by EDC. Any number of vehicles, trailers, and/or barriers may be included. Simultaneous collisions between any numbers of objects are allowed. The original SMAC collision model has also been extended significantly and now includes different stiffness parameters for each side of the vehicle, as well as direct support for barrier collisions. It also provides features for occupant crash simulations (see EDHIS and Articulated Total Body (GATB) in the succeeding text). While EDSMAC4 has been extended and revalidated for use in HVE to account for 3D terrain, EDSMAC4 is essentially a 2D reconstruction program.
- *EDVSM* is a simulation program module for the true 3D dynamic response of a motor vehicle to driver inputs and 3D road conditions. Based on the HVOSM-VD2 model developed at Calspan, EDVSM incorporates enhancements, including a tire–terrain interaction model and a tire blowout model. The user has to enter initial position, velocity, and driver controls. The program will calculate the 3D vehicle kinematics, the tire forces and moments, the suspension forces and deflections, the aerodynamic forces, and the drive and brake system torques. Vehicle body versus terrain interaction is also modeled by calculating the force between the vehicles 3D mesh (see Section 41.5.1.2) and the terrain mesh. With EDVSM, a vehicle may be simulated driving on virtually any surface under any conditions, including airborne vehicles and complete vehicle rollover.
- *EDVDS* is a simulation program module for the 3D dynamic response of a commercial vehicle to driver inputs and 3D road conditions. A tow vehicle and up to three trailers may be simulated. EDVDS is an extended version of the phase 4 program developed at UMTRI, and its features are comparable to those of EDVSM.
- *SIMON* is another simulation program module for the 3D dynamic response of one or more vehicles to driver inputs, intervehicle collision(s), and factors related to the environment. SIMON is a newly developed simulation module, using a new, general purpose 3D vehicle

dynamics engine (see [53,54]). The dynamics engine allows a sprung mass with six DOFs and multiple axles with up to five DOFs per axle. The user assigns initial position, velocity, and driver controls for each vehicle. Trailers and dollies are automatically positioned behind a tow vehicle according to connection compatibility and location. SIMON then predicts the motion of each vehicle's sprung mass and the motions of the unsprung masses, suspension dynamics, tire dynamics, and current brake system parameters at each wheel. Vehicle body damage from a collision simulation is also calculated and displayed using HVE's 3D viewers. SIMON was specifically designed to take advantage of the advanced features available in the HVE simulation environment.

- *EDHIS* is a simulation program module for the 3D analysis of the response of a human occupant during a motor vehicle collision. EDHIS is based on the HSRI 3D model developed at UMTRI and includes several extensions and refinements. It employs airbag and belt (torso and lap) models suitable to study issues related to restraint system effectiveness. EDHIS is a 12-DOF analysis, dividing the human body into three masses (head, torso, and lower extremities) and two joints (neck and hips). The body is visualized using the HVE 15-segment/14-joint human model.
- *GATB* is a 3D computer model based on the computer program ATB. The GATB model may be used to study occupants or pedestrians involved in motor vehicle collisions. GATB represents the human body with 15 segments and 14 joints. The GATB model has 48 DOFs for each human. Each of the 15 mass segments can move and interact with the environment. GATB computes human kinematics, joint angles and torques, contact forces between the human and contact panels attached to the vehicle, and contact forces between humans or between a single human's limbs. By using GATB, up to four human models can be placed inside a car during a crash to predict how the humans would move, what part of the interior of a car they would hit, or how an ejection could occur. GATB can also simulate a human body being struck by a vehicle's exterior to study pedestrian impacts.

41.5.1.2 Advanced Features in HVE

- *DyMESH*: A computer model for 3D dynamic simulation of motor vehicle crashes. DyMESH employs methods from FE technology for collision detection and stress–strain relationships for force calculation. DyMESH uses a 3D vehicle mesh with mechanical properties as input and produces vehicle-fixed collision forces and moments as output. The output from DyMESH is then used by the vehicle simulation module along with suspension forces, aerodynamic forces, and intervehicle connection forces to produce the total vehicle-fixed forces and moments acting on the vehicle at each time step. Whereas a full FE analysis requires several hours or even days, a simulation using DyMESH is completed in a few minutes.

  DyMESH is useful for all collision simulations and is especially useful for underride or any crash where 3D collision dynamics are present. Detailed validation results can be found in [50,51].
- *HVE driver model*: This is a closed-loop driver model, allowing users to define an attempted maneuver and have the simulation determine the required steering inputs to make the vehicle follow the defined path. Typical examples include single and double lane-change maneuvers. The HVE driver model may also be used to study the effects of driver fatigue and intoxication.
- *HVE steer* DOF *model*: It allows users to study how hands-off steering affects vehicle trajectory. The steering inputs used in the simulation are calculated from the interaction of the vehicle's tires travelling over and interacting with the 3D terrain model defined by the user. An obvious application of the steer DOF model is the postcollision motion of a vehicle.
- *HVE tire blowout model*: A model that provides detailed information regarding the transient nature of a vehicle's response to a tire blowout that occurs during any maneuver.

- *Other models*: The *ABS model* simulates the ABS system on an actual vehicle. The *Brake Designer* includes the capability to study brake fade characteristics. The *Radial Spring Model* allows for simulating tires mounting curbs and rolling over potholes. The *Sidewall Impact Model* allows for a rigorous modeling of the forces applied to the sidewall of a tire such as occurs in a curb-tripped rollover. The *Soft Soil Model* allows for calculating tire drag for a vehicle that has left the roadway and is travelling through soft dirt.

41.5.2 PC Crash

PC Crash and its constricted version PC Crash 2D are both products of Dr. Steffan Datentechnik (DSD), located in Linz, Austria. The programs are distributed in North America by MEA Forensic Engineers & Scientists, located in Richmond, Canada, and by several distributors in Europe (see the webpage [55]). The following overview and product description reflects the product's development status when this book was published and is based on information and material available from the product websites [55,56].

PC Crash is a MS-Windows-based 3D collision and trajectory simulation tool that enables the analysis of a wide variety of motor vehicle collisions and other incidents. Collisions with up to 32 vehicles can be simulated in 2D and also in 3D. Car-to-car accidents, car-to-motorcycle accidents, car-to-pedestrian accidents, occupant movement, and also rollover can be investigated. Currently, PC Crash is maybe the most popular accident reconstruction program with more than 2700 installations in countries around the world.

The basic models and features of PC Crash are summarized briefly:

- *Impact models*: PC Crash uses the Kudlich–Slibar impact model [22,45] as the primary tool for collision analysis. Primary and secondary impacts are detected automatically. Secondary impacts can have different properties than specified for the primary impact. A very useful feature is the collision optimizer. One may define an objective function to quantify the quality of any given solution. PC Crash will then vary the quantities specified, including vehicle speeds, impact properties, and vehicle positions. The collision optimizer will find a solution that minimizes the objective function.

  Another impact model available, and especially useful with 3D models, is a deformable mesh model. By specifying the mesh force/deflection characteristics, PC Crash will determine contact forces and deformations over time, modifying the vehicle trajectory accordingly. Meshes can interact with other deformable meshes, rigid objects, or the ground. This model is useful when vehicle deformations may significantly affect subsequent vehicle interactions.

  For impacts with sustained, low-level contact or impacts with a significant vertical component, a so-called ellipsoid model is available. The vehicle is modeled as a set of compressible ellipsoidal elements. This is the default rollover model and can be used for vehicle-to-vehicle collisions as well.
- *Trajectory models*: Full 3D rigid-body dynamics of vehicles is implemented. Tire forces are generated by either a simple tire model or the more sophisticated TMEasy model. Vehicle features like ABS and also stability control can be accounted for.
- *Trailer models*: Fully customizable trailers can be attached to any vehicle. All the same model features apply to trailers as to normal vehicles. Forces are transmitted at the hitch to satisfy kinematic constraints with connected vehicles. It is possible to combine multiple trailers for B-trains and even more complex combinations. Hitches have user-definable properties, including rotational stiffness along each of three axes and maximum allowable hitch force before failure.
- *Terrain models*: Terrain models may be imported as 3D DFX graphics or designed by built-in 2D and 3D drawing tools. Parametric road shape specification is available.
- *MADYMO:* The optional MADYMO module allows FE modeling of vehicle occupant motions and loading (see also [57]). One can use the simulation already completed in PC

Crash, and then the MADYMO module will simulate the motions of a Hybrid-3 crash test dummy. Different types of seat belts can be specified, including a seat belt pretensioner when a lap and torso belt is used. Two different airbags (driver's and right front) can be specified. The results of a MADYMO simulation can be viewed as a 3D animation or in the form of diagrams.

- *Other features*: Simulation of a movable load is possible by using the multibody system of PC Crash. Its influence on the vehicle's driving behavior can be investigated. Additional features include the capability of defining and calculating brake force distributions between front and rear axle, a pedestrian and motorcycle model (multibody model) where the real vehicle shape is used in the calculation, scalable pedestrians by anthropometrical data, a preprocessor to define and configure multibody systems (pedestrians, motorcycles with drivers, etc.), the calculation of tire contact marks, and energy equivalent speed (EES) calculation for Crash-3 models.

PC Crash 2D is a constricted version of the full program, has no 3D capabilities, and allows only two vehicles and no trailers. Multibody models are not available in this version either.

41.5.3 VCRware

VCRware 2.0 is a product of Brach Engineering, located in Granger, Indiana. The following overview and product description reflect the product's development status when this book was published and is largely based on information and material available from the company's website [58].

VCRware is a suite of altogether eleven spreadsheets, based on MS-Excel, and allows the reconstructionist to analyze and reconstruct most cases of vehicular accidents. The documentation of this software suite and the theoretical basis for each of the programs is covered in detail in [7].

The Excel spreadsheet environment provides analysis functions that allow the user to determine the value of an input parameter to achieve a specified or desired value of some spreadsheet output without the need for iterative evaluation. For instance, the user can easily determine in one step the drag factor needed to stop a vehicle within a given distance at a certain initial speed. Also, complex problems such as maximization and minimization with multiple variables can be handled and enable the user to automate the time-consuming iterative process typically necessary to perform many reconstruction tasks. In addition, statistical tools for the analysis of the uncertainty in calculations using Monte Carlo techniques can be conveniently used in conjunction with VCRware.

With regard to vehicle dynamics, the most important spreadsheets are as follows:

- *Articulated vehicle impact*: For two vehicles, which may be articulated or not, with known physical parameters, orientations in a reference frame, and initial velocity components, this spreadsheet uses planar impact mechanics (see [59]) to calculate the final velocity components after the impact and also delta-V values and the energy loss.
- *Critical speed formula*: Determines the (critical) speed of a single vehicle making yaw marks over a circular path following a sudden steer maneuver for a given frictional drag coefficient and measured curve coordinates (see [7,60]).
- *Planar impact mechanics analysis*: A very useful spreadsheet to calculate the final velocity components, the delta-V values, and the energy loss from planar impact mechanics, for two vehicles with known physical parameters, orientations, and collision centers. See [7] for a comprehensive documentation of the underlying model.
- *Low-speed vehicle impact analysis*: Analysis of a low-speed front-to-front or front-to-rear collision of two vehicles using point-mass collision mechanics including the collision coefficient of restitution and vehicle roadway drag, when appropriate (see also [61]).
- *Crush stiffness coefficients from vehicle to barrier test*: This spreadsheet determines crush stiffness coefficients using data from a vehicle-to-barrier test.

- *Energy loss and speed change from CRASH3*: A spreadsheet to calculate the energy loss and delta-V values using the CRASH3 algorithm for a two-vehicle collision.
- *Vehicle–pedestrian collision analysis*: The user can calculate the times, velocities, pedestrian throw distance, and vehicle braking distance, corresponding to the various events and secondary events of a vehicle–pedestrian (or bicycle rider, when appropriate) collision. A reference covering basics of this kind of accidents is [62].
- *Vehicle and semitrailer dynamics program*: Calculates the planar motion of a vehicle alone or a vehicle pulling a semitrailer as function of time. Driver control modes include locked-wheel braking, lane-change maneuver, or arbitrary steering. Individually, locked wheels, drive train drag, and uniform acceleration can be simulated.

41.5.4 Other Software

As mentioned already, there are quite a large number of software products available to assist in accident analysis and reconstruction. Most of them have dedicated webpages and can be easily found by a search engine using suitable search words. In addition to that, Ref. [8] may be helpful since it discusses about seven more software packages to more or less extent.

41.6 SPECIAL TOPICS IN ACCIDENT ANALYSIS

Due to the variety of motorized and nonmotorized vehicles, pedestrians, and even animals that populate roads, there are quite a number of accident scenarios other than car to car collisions. The following sections will give a brief introduction to the peculiarities of some of these accidents and their reconstruction.

41.6.1 Motorcycle Accidents

From the statistics, it is known that motorcycle accidents do contribute significantly to the numbers, especially when it comes to severely injured and killed participants in accidents. One of the reasons for this is the fact that PTWs are among those vehicles with high potential for acceleration and velocity on one hand but with very limited passive safety for the rider on the other hand.

Also single-track vehicles are unique among other vehicles since they are inherently unstable with respect to their upright position at rest and at most speeds (see Chapter 35). This may be of importance with certain single vehicle accidents, where vehicle instability is the cause of an accident. There is rarely evidence for such kind of incident other than observation by the rider and, in severe cases, sometimes skid marks on the road.

The precrash phase of a PTW has to be analyzed differently compared to regular motor vehicles since PTWs still mostly have two independently operating braking systems [63]. Even with skid marks at the accident scene due to a locked wheel (mostly the rear wheel), the braking deceleration could have been within a rather wide range and may be difficult to estimate [64]. Furthermore, the PTW rider may have the choice to make an evasive lane-change maneuver instead of trying to stop the vehicle by just braking.

The crash of a PTW against another vehicle may look very different depending on which part of the opponent vehicle is struck by the PTW and also depending on the mass ratio of both vehicles (see, e.g., [65]). Permanent deformations at the front of a PTW are used to reconstruct the impact velocity, but they do not give very reliable results. Impact mechanics may be applied to analyze the collision, but attention has to be paid to the rider, who may interact with the handlebar, with the other vehicle or even may separate early from the PTW.

Rest positions of a PTW and the rider are very useful pieces of information to reconstruct the post-collision speeds and trajectories. Deceleration parameters need to be chosen carefully, since the kind of

PTW and type of ground have a significant influence on the sliding friction. The post-crash trajectories of both PTW and rider also may be travelled to a large part in the air and therefore with almost no friction.

A comprehensive discussion of many aspects of motorcycle accidents and extensive references are found in the books [9,10,12,66], the latter including many results from crash tests.

41.6.2 Accidents Involving Pedestrians

In most cases of accidents involving pedestrians, the dynamics of the hitting vehicle does not play a major role in the accident scenario. Since pedestrians and vehicles are normally using separate areas of the streets, conflicting situations almost only occur when pedestrians enter the areas used by vehicles. This may legally happen on a crosswalk, and the vehicle driver may run a red light, or it may illegally happen when, for example, a pedestrian attempts to cross a road where he or she is not supposed doing this, enters the lane of the approaching vehicle, and is being hit.

The task of reconstructing a vehicle–pedestrian accident practically always starts with the questions of visibility and recognition of the pedestrian by the vehicle driver. The main purpose is to locate the position and the point in time when it seems reasonable that the driver could initiate an action to avoid the accident. Of course, the same considerations apply to the pedestrian if he or she has not the right of way. Mainly, this question is a matter of perception, recognition, and reaction time, influenced by human factors like health and mental condition, medication, alcohol or drug usage, and environmental factors like daylight and weather conditions.

Mostly, the motion of the vehicle can be assumed to be of constant speed or of decelerating at a constant rate, in the precrash phase, and thereafter until the rest position is reached. Due to the large mass ratio of a vehicle in relation to a pedestrian, the collision causes only a minor or even negligible change of speed for the vehicle. Speed and distance versus time can be calculated for the vehicle, and for the pedestrian as well, from kinematic formulas as presented in Section 41.3.1.

The most challenging part in vehicle–pedestrian accidents is the reconstruction of the collision speed of the vehicle. If the position of the initial pedestrian contact and the rest position of the vehicle are known from some evidence, then this postimpact distance can be used to calculate the collision speed for the vehicle. If this is not possible because the vehicle did not stop or the deceleration cannot be determined, then one may try to use the throw distance of the pedestrian instead. There are empirical results available from tests with dummies and with cadavers but also equations derived from more or less simple mechanical models. The motion of a pedestrian, who is stroke by a vehicle, is quite complex, starting with one or more contacts with the vehicle, then a phase of free flight follows, and the final rest position is reached after some rolling and/or sliding on the ground. Depending on the goals of the accident reconstruction, it might be sufficient to use approximate models as summarized in [7], but occasionally sophisticated multibody dynamics software like MADYMO [57] may be needed to capture the complex interaction during the collision between the vehicle and the pedestrian.

A detailed discussion of the various aspects of vehicle–pedestrian accidents and related references are found in the books [7,9,10,17,32].

REFERENCES

1. I. Fallon and D. O'Neill, The world's first automobile fatality, *Accident Analysis & Prevention* 37 (2005), pp. 601–603.
2. N.N., *World's First Automobile Accident,* Ohio History Central, revision April 28, 2013, www.ohiohistorycentral.org.
3. N.N., Pillars and posts, Autocar 137 (1972) Haymarket Motoring Publications Ltd.
4. World Health Organization, *Global Status Report on Road Safety: Time for Action,* World Health Organization, Geneva, Switzerland (2009).
5. I. Parry and P. Fidler, *The Simulation of the Fatal Crash Involving Diana, Princess of Wales and Implications for the Investigation of Loss of Control Incidents,* in *17th EVU Annual Congress,* Nice, France, November 6–8, 2008, pp. 151–164.

6. R. Nader, *Unsafe at Any Speed*, *The Designed-In Dangers of The American Automobile,* Grossman Publishers, New York (1965).
7. R.M. Brach and R.M. Brach, *Vehicle Accident Analysis and Reconstruction Methods,* SAE, Warrendale, PA (2005).
8. R.V. Dukkipati, J. Pang, M.S. Qatu, G.S. Chen, and Z. Shuguang, *Road Vehicle Dynamics,* SAE, Warrendale, PA (2008).
9. H. Burg and A. Moser (eds.), *Handbuch Verkehrsunfallrekonstruktion,* Friedr. Vieweg & Sohn, Wiesbaden, Germany (2007).
10. W. Hugemann (ed.), *Unfallrekonstruktion,* Verlag autorenteam, Münster, Germany (2007).
11. R. Fucik, F. Hartl, H. Schlosser, and B. Wielke (eds.), *Handbuch des Verkehrsunfalls,* 2nd edn. MANZ, Vienna, Austria (2008).
12. K.S. Obenski, P.F. Hill, E.S. Shapiro, and J.C. Debes, *Motorcycle Accident Reconstruction and Litigation,* 4th ed. Lawyers & Judges Publ. Comp., Tucson, AZ (2007).
13. R.F. Sutphen and R.W. Varner, *Commercial Vehicle Accident Reconstruction and Investigation,* Lawyers & Judges Publ. Comp., Tucson, AZ (2008).
14. R.S. Hickman and P.E. Hill, *Bus and Recreational Vehicle Accident Reconstruction and Litigation,* Lawyers & Judges Publ. Comp., Tucson, AZ (2000).
15. R.E. Dewar and P.L. Olson, *Human Factors in Traffic Safety,* Lawyers & Judges Publ. Comp., Tucson, AZ (2002).
16. A.T. Welford, *Reaction Times,* Academic Press, London, U.K., 1980.
17. P.L. Olson, *Forensic Aspects of Driver Perception and Response,* Lawyers & Judges Publ. Comp., Tucson, AZ (1996).
18. Michael S. Varat, *Crash Reconstruction Research: 20 Years of Progress (1988–2007),* PT-138 SAE, Warrendale, PA (2008).
19. R. Fay, R. Robinette, J. Scott, and D. Deering, *Using Event Data Recorders in Collision Reconstruction,* Paper 2002-01-0535, SAE, Warrendale, PA (2002).
20. B. Brogliato, *Nonsmooth Mechanics,* 2nd edn. Springer, New York (1999).
21. W.J. Stronge, *Impact Mechanics,* Cambridge University Press, Cambridge, U.K. (2000).
22. G. Kasanický, P. Kohút, and M. Lukášik, *Impact Dynamics Theory for the Analysis and Simulation of Collisions,* Žilinská Univerzita, Žilina, Slovak Rep. (2004).
23. M. Huang, *Vehicle Crash Mechanics,* R-284 SAE, Warrendale, PA (2002).
24. W. Kalthoff, *The Coefficient of Restitution on Rear-End Collisions. New Findings,* in *17th EVU Annual Congress,* Nice, France, November 6–8, 2008, pp. 283–288.
25. K.L. Campbell, *An Energy Basis for Collision Severity,* Paper 74565, SAE, Warrendale, PA (1974).
26. N.J. Carpenter, *Stiffness and Crush Energy Analysis for Vehicle Collision and Its Relationship to Barrier Equivalent Velocity (Bev),* Paper 2001-01-0500, SAE, Warrendale, PA (2001).
27. A. Cesarani, D. Alpini, R. Boniver, C. Claussen, P. Gagey, L. Magnusson, and L. Odkvist (eds.), *Whiplash Injuries,* SAE, Warrendale, PA (1996).
28. J.A. Pike, *Forensic Biomechanics,* R-379 SAE, Warrendale, PA (2008).
29. J.A. Pike, *Neck Injury Biomechanics,* PT-141 SAE, Warrendale, PA (2009).
30. D.C. Viano, *Role of the Seat in Rear Crash Safety,* R-317 SAE, Warrendale, PA (2002).
31. N.N., *2009 Occupant Protection and Crashworthiness Technology Collection on CD-ROM,* OP2009 SAE, Warrendale, PA (2009).
32. J.J. Eubanks and P.F. Hill, *Pedestrian Accident Reconstruction and Litigation,* 3rd ed. Lawyers & Judges Publ. Comp., Tucson, AZ (2009).
33. R.H. Barnard (ed.), *Road Vehicle Aerodynamic Design,* 2nd ed. Mechaero Publishing, St Albans, U.K. (2001).
34. P. Lugner and M. Plöchl (eds.), *Tyre Models for Vehicle Dynamics Analysis,* Taylor & Francis Group, London, U.K. (2005).
35. P. Lugner and M. Plöchl (eds.), *Tyre Model Performance Test,* Taylor & Francis Group, London, U.K. (2007).
36. R.M. Brach and R.M. Brach, *Tire Models for Vehicle Dynamic Simulation and Accident Reconstruction,* Paper 2009-01-0102, SAE, Warrendale, PA (2009).
37. H.B. Pacejka, *Tyre and Vehicle Dynamics,* 2nd edn., Butterworth Heinemann, Oxford, U.K. (2005).
38. E. Bakker, L. Nyborg, and H.B. Pacejka, *Tyre Modelling for Use in Vehicle Dynamics Studies,* Paper 870421, SAE, Warrendale, PA (1987).
39. R.M. Brach and R.M. Brach, *Tire Forces: Modeling the Combined Braking and Steering Forces,* Paper 2000-01-0357, SAE, Warrendale, PA (2000).

40. W. Hirschberg, G. Rill, and H. Weinfurter, User-appropriate tyre-modelling for vehicle dynamics in standard and limit situations, *Vehicle System Dynamics* 38 (2003), pp. 103–125.
41. W. Hirschberg, G. Rill, and H. Weinfurter, Tire model TMeasy, *Vehicle System Dynamics* 45 (2007), pp. 101–119.
42. C.C. MacAdam, P.S. Fancher, G.T. Hu, and T.D. Gillespie, *A Computerized Model For Simulating the Braking and Steering Dynamics of Trucks, Tractor-Semitrailers, Doubles, and Triple Combinations,* UM-HSRI 80-58, The University of Michigan, Ann Arbor, MI, 1980.
43. E. Fiala, *Seitenkräfte am rollenden Luftreifen,* VDI-Zeitschrift 96 (1954).
44. R.M. Brach, *Mechanical Impact Dynamics,* John Wiley & Sons, New York, 1991.
45. H. Kudlich, *Beitrag zur Mechanik des Kraftfahrzeug-Verkehrsunfall,* PhD thesis., Vienna University of Technology, Vienna, Austria (1966).
46. P.H. Cheng, M.J. Sens, J.F. Weichel, and D.A. Guenther, *An Overview of the Evolution of Computer Assisted Motor Vehicle Accident Reconstruction,* Paper 871991, SAE, Warrendale, PA (1987).
47. A.K. Prasad, *CRASH3 Damage Algorithm Reformulation for Front and Rear Collisions,* Paper 900098, SAE, Warrendale, PA (1990).
48. T. Noga and T. Oppenheimer, *CRASH3 User's Guide and Technical Manual,* DOT Report HS 805 732, NHTSA, Washington, DC (1981).
49. F. Zeidler, H.H. Schreier, and R. Stadelmann, *Accident Research und Accident Reconstruction by the EES-Accident Reconstruction Method,* Paper 850256, SAE, Warrendale, PA (1985).
50. A.R. York and T.D. Day, *The DyMesh Method for Three-Dimensional Multi-Vehicle Collision Simulation,* Paper 1999-01-0104, SAE, Warrendale, PA (1999).
51. A.R. York and T.D. Day, *Validation of DyMESH for Vehicles vs. Barrier Collisions,* Paper 2000-01-0844, SAE, Warrendale, PA (2000).
52. Engineering Dynamics Corporation, Beaverton, OR, http://www.edccorp.com/, visited May 10, 2013.
53. T.D. Day, S.G. Roberts, and A.R. York, *Simon: A New Vehicle Simulation Model for Vehicle Design and Safety Research,* Paper 2001-01-0503, SAE, Warrendale, PA (2001).
54. T.D. Day, *Validation of the Simon Model for Vehicle Handling and Collision Simulation—Comparison of Results With Experiments and Other Models,* Paper 2004-01-1207, SAE, Warrendale, PA (2004).
55. Dr. Steffan Datentechnik, Linz, Austria, http://www.dsd.at/, visited May 10, 2013.
56. MEA Forensic Engineers & Scientists, Richmond, BC, http://www.pc-crash.com, visited May 10, 2013.
57. TASS, Rijswijk, The Netherlands, http://www.tass-safe.com, visited May 10, 2013.
58. Brach Engineering, Granger, IN, https://www.brachengineering.com, visited May 10, 2013.
59. R.M. Brach, *Analysis of Collisions Involving Articulated Vehicles,* Paper 2007-01-0735, SAE, Warrendale, PA (2007).
60. R.M. Brach, *An Analytical Assessment of the Critical Speed Formula,* Paper 970957, SAE, Warrendale, PA (1997).
61. R.M. Brach, *Modeling of Low-Speed, Front-to-Rear Vehicle Impacts,* Paper 2003-01-0491, SAE, Warrendale, PA (2003).
62. I. Han and R.M. Brach, *Throw Model for Frontal Pedestrian Collisons,* Paper 2001-01-0898, SAE, Warrendale, PA (2001).
63. H. Ecker, J. Wassermann, R. Ruspekhofer, G. Hauer, and M. Winkelbauer, *Brake Reaction Times of Motorcycle Riders,* in *Proc. of Int. Motorcycle Safety Conference,* Orlando, FL, March, 1–4 2001.
64. H. Ecker, J. Wassermann, G. Hauer, R. Ruspekhofer, and M. Grill, *Braking Deceleration of Motorcycle Riders,* in *Proc. of Int. Motorcycle Safety Conference,* Orlando, FL, March 1–4, 2001.
65. K.S. Adamson, P. Alexander, C.I.B. III, J.R. Kinney, R. Aronberg, J. McManus, G.C. Anderson, D.W. Sallmann, G. Johnson, and A.G. Fonda, *Seventeen Motorcycle Crash Tests Into Vehicles and a Barrier,* Paper 2002-01-0551, SAE, Warrendale, PA (2002).
66. G. Kasanický, P. Kohút, and J. Priester, *Analysis of Single-Track Vehicle Accidents,* Žilinská Univerzita, Žilina, Slovak Rep. (2003).

42 Automotive Structural Crashworthiness and Occupant Protection

Jorge Ambrósio

CONTENTS

42.1 Introduction 1611
42.2 Measuring Occupant Protection 1616
42.2.1 Injury Criteria for Different Body Parts 1616
42.2.1.1 Head Injuries 1617
42.2.1.2 Spinal Injuries 1617
42.2.1.3 Thoracic Injuries 1618
42.2.1.4 Abdominal Injuries 1619
42.2.1.5 Pelvic Injuries 1619
42.2.1.6 Lower Extremity Injuries 1619
42.2.1.7 Relation between Injury Criteria and Occupant Protection 1620
42.2.2 Regulations for Automotive Crash Testing 1621
42.2.3 Anthropometric Testing Devices: Dummies 1626
42.2.4 Impact Barriers 1628
42.3 Structural Features and Systems for Passive Safety 1628
42.3.1 Vehicle Structure for Crashworthiness 1630
42.3.2 Vehicle Interior Protection Systems 1632
42.4 Numerical Methods in Vehicle Passive Safety 1633
42.5 Multibody Dynamics-Based Analysis Tools: Basics 1635
42.5.1 Contact Detection and Contact Models 1636
42.5.2 Plastic Hinge Approach to Structural Deformations 1639
42.6 Development of a Vehicle Model for Front Impact 1641
42.7 Development of Side Impact Protection Systems 1649
42.8 Conclusions and Trends in Future Developments 1657
Acknowledgments 1658
References 1658

42.1 INTRODUCTION

The body of the vehicle is primarily designed to carry the occupants and cargo, the vehicle powertrain, and the chassis, and, only after, to mitigate human injury of occupants or other road users. Vehicle structures and systems experienced a dramatic change in their configuration during the past 30 years to accommodate the requirements for an improved protection of passengers and cargo in the event of a crash. The research area of vehicle passive safety, which deals with the study of structural components of the vehicle and the different types of occupant restraint and protection systems, addresses all events from the moment a crash starts until it is over. During this time period, it is necessary to control the vehicle

deformation so that it progresses from the exterior to the interior, maintaining the occupants' survival space as intact as possible, to shape the profiles of the energy absorption. By exploiting properly structural components and vehicle safety systems, the occupants have not to be exposed to high accelerations or impact forces, and all the occupant protection systems, such as seat belts, airbags, seat positions, and configurations, that control the occupant kinematics have to be managed. The aim of this chapter is to provide some historical background on the developments in automotive vehicle passive safety, to present requirements and regulations for their design and approval, and to propose and demonstrate some of the numerical tools used in the design and analysis of vehicles for crashworthiness.

Since automotive vehicles have to share the road among themselves and with other users, the potential for accidents exists, as highlighted in the advertisement presented in Figure 42.1, regardless of all measures taken to avoid collisions. Vehicle passive safety drew the interest of the different actors on the automotive-related fields since the early days of automobiles. The protection of occupants subjected to impact loading or to high acceleration started with the pioneering work of Hugh DeHaven, an air crash survivor often referred as the father of crashworthiness, who developed the initial studies on injury biomechanics and published the first work on this topic [1]. During the 1950s, John Paul Stapp, colonel of the U.S. air force, developed an experimental program to evaluate the human tolerance to extreme accelerations in order to establish the limits for human tolerance. Colonel Stapp [2] is well known not only for the testing program that he developed but also because he performed some of the experiments on himself, being the first human that willingly decelerated from about 1000 km/h to a full stop in about 1.4 s using the setup shown in Figure 42.2. He has his

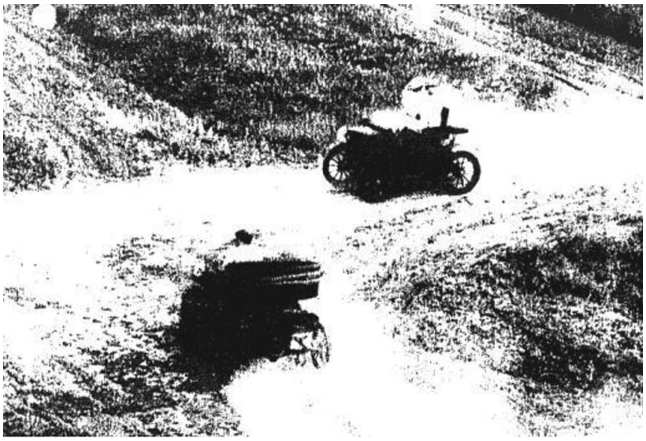


FIGURE 42.1 There were only two cars in Ohio in 1895. Guess what happened? (From an advertisement.)



FIGURE 42.2 Rocket-propelled sled used by Colonel Stapp in the experimental program on human tolerance to extreme accelerations.

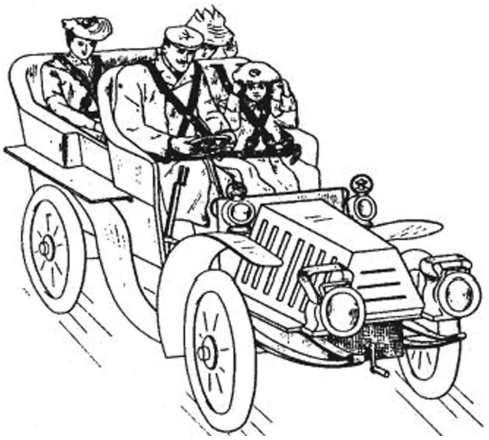


FIGURE 42.3 First concept for the seat belt proposed by Lebau as presented in his patent. (From *Proceedings of the 8th Stapp Car Crash Conference*, Detroit, MI, 1968.)

name given to a series of conferences, Stapp Car Crash Conference, that is being run annually since 1961 [3]. Many other individuals and groups contributed during the years after the World War II for studies in impact biomechanics that led to the developments that made possible the occupant protection systems that are implemented in commercial vehicles today. For a review on these developments, the interested reader is referred to a review by Mackay [4] or to the work of Schmitt et al. [5].

Although the first works aimed at understanding the biomechanics of injury have been developed during the mid-twentieth century, the first proposal for restraint systems in automobiles dates from 1903 and consists in the seat belt patent by Lebau [6], shown in Figure 42.3. However, seat belts have not been used in road vehicles until after the World War II. The three-point seat belt, a standard used today in all automobiles, was invented in 1959 by the Volvo engineer Nils Bohlin. Based on findings on human tolerance to impact and high accelerations and on the available technology, the mandatory seat belts usage, for driver and front passengers, was first introduced in several countries: from Oceania, in the early 1970s; in most European countries by the mid-1970s; and later in the United States by the mid-1980s [7]. The mandatory use of seat belt for rear passengers was introduced by the mid-1980s and early 1990s [8].

Another of the passive safety systems widely used in automotive vehicles is the airbag, its invention credited, independently, to the German Walter Linderer and to the North American John Hetrick who in 1951 registered the first patents for its early designs. Although the invention of the airbag resulted from simple observation, without any use of the knowledge on injury biomechanics, the principles involved in this device were further explored by General Motors, Ford, and Mercedes-Benz, among others, during the 1960s when proper sensing equipment and the ignition of a gas generator propellant to inflate the airbag were designed. From the first commercial vehicle using airbag for the driver, the 1973 Oldsmobile Toronado, passing by the 1980 S-class Mercedes-Benz until current vehicles, the airbag's acceptance as being one of the most important passive safety systems in a car grew continuously. Since 1997, all new cars produced in the United States are required to have front airbags for the driver and passenger. Almost all vehicles produced today worldwide have, besides the mandatory front airbags, also side airbags and, in many of them, thorax and knee airbags. A common setup in many current commercial vehicles is similar to that presented in Figure 42.4.

A perceived requirement to protect the occupant, until the 1960s, consisted in having stiff automobile bodies that would not deform during a crash. This construction concept was first disputed by Barény [9], an engineer for Mercedes-Benz, who proposed in 1952 that the vehicle strength should increase from the exterior of the car to its interior so that the progressive crushing would be responsible



FIGURE 42.4 Airbags used in current cars for occupant protection in frontal and side impact.

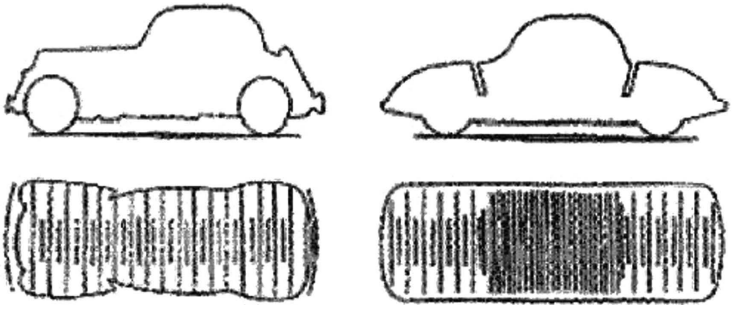


FIGURE 42.5 The crumple zone concept proposed by Barény. (From Barényi, B., Kraftfahrzeug, Insbesondere zur Beförderung von Personen, Deutsches Patentamt Nr. 854157, 30, 1952.) The density of the transversal lines is a representation of the vehicle body stiffness.

for energy absorption during a crash, as shown in Figure 42.5. Although the vehicle body is designed to carry the main vehicle components and the occupants, the concept of a crumple zone has been refined and widely applied to all vehicles in production today being the body in white of the vehicle responsible for absorbing over one-quarter of the total kinetic energy during a crash. Still it is important that the survivable cell of the vehicle is maintained basically intact not only to avoid intrusions that would decrease this space but also to ensure that the different safety systems are not compromised.

Having established most of the basic principles and systems in which the automotive vehicle occupant protection is based, the emphasis is put on the implementation of the structural arrangements and systems where they matter more. The analysis of the accident statistics serves as the basis to define the priorities for the development, implementation, and regulation of passive safety in road vehicles. Figure 42.6 shows a typical distribution of car accidents in Europe [10–12] and the incidence of fatalities for each collision category [12]. For instance, front accidents account for about 48% of the total accidents and for 40% of the accidents with fatalities. Therefore, it is no surprise that the front structure of the vehicles and occupant protection systems for front crash were the first deserving attention.

The crash–injury recording systems existing in many countries are generally based in basic police reports linked with hospital records. Although not always reliable, the situation has improved dramatically over the past years and the accident statistics provide the most important source of information for the decision-making policies in passive safety [4]. However, in what minor injuries are concerned, there is still a strong underreporting and improved links between hospital and police reporting is needed.

Being an important part of the evaluation of injury in traffic accidents, the medical reporting has to be based on some form of quantification of injury that can be understood by medical personnel,

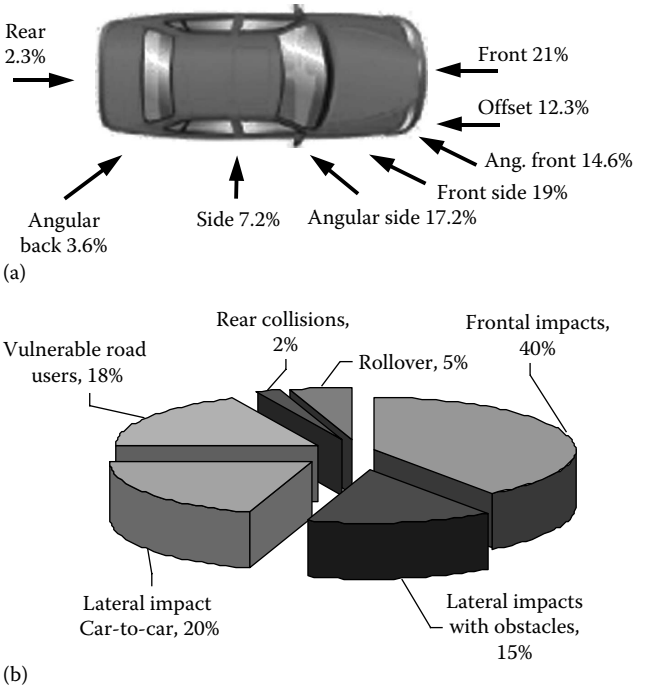


FIGURE 42.6 Accidentology analysis: (a) distribution of vehicle collision locations (From Seiffert, U. and Wech, L., *Automotive Safety Handbook*, SAE International, Warrendale, PA, 2003.) and (b) fatality rate for the individual collision categories. (From Klanner, W., Status report on future development of EuroNCAP programme, *Proceedings of the 17th ESV Conference*, Amsterdam, the Netherlands, 2001.)

engineers, and regulators alike. Injury scaling is defined as the numerical classification of the type and severity of an injury. The schemes proposed for ranking and quantifying injuries are either anatomic, physiologic, or impairment, disability, and societal loss scales [13]. Among the anatomic scales, the Abbreviated Injury Scale (AIS), published in 1971 and updated several times afterward [5,14], provides a classification of injuries resulting from traffic accidents based on medical diagnosis. The AIS classification is described in Table 42.1 together with the fatality rate associated to each score [13]. Table 42.2 presents examples of the types of injury sustained to achieve each score.

It should be noted that the AIS is not a linear scale and that it is not accepted universally as a score for injury. One of the main criticisms is that the AIS does not access multiple injuries.

TABLE 42.1
AIS and Fatality Rate Associated to Each Score

AIS	Severity Code	Fatality Rate (Range%)
0	No injury	0.0
1	Minor	0.0
2	Moderate	0.1–0.4
3	Serious	0.8–2.1
4	Severe	7.9–10.6
5	Critical	53.1–58.4
6	Not survivable	100

TABLE 42.2
Examples of AIS Scores per Body Region

	Body Region				
AIS	Head	Thorax	Abdomen and Pelvic Contents	Spine	Extremities and Bony Pelvis
0					
1	Headache or dizziness	Single rib fracture	Abdominal wall: superficial laceration	Acute strain (no fracture or dislocation)	Toe fracture
2	Unconscious less than 1 h; linear fracture	2–3 rib fracture; sternum fracture	Spleen kidney or liver: laceration or contusion	Minor fracture without any cord involvement	Tibia, pelvis, or patella: simple fracture
3	Unconscious 1–6 h; depressed fracture	≥4 rib fracture; 2–3 rib fracture with hemo- or pneumothorax	Spleen or kidney: major laceration	Ruptured disk with nerve root damage	Knee dislocation; femur fracture
4	Unconscious 6–24 h; open fracture	≥4 rib fracture with hemo- or pneumothorax; flail chest	Liver: major laceration	Incomplete cord syndrome	Amputation or crush above knee; pelvic crush
5	Unconscious more than 24 h; large hematoma	Aorta laceration (partial transection)	Kidney, liver, or colon rupture	Quadriplegia	Pelvic crush (open)
6	—	—	—	—	—

The Injury Severity Score [14], based on the AIS for the different body regions, is preferred by some authorities, while others chose different classifications. However, it is important to note also that when a relation between the injury scales and the injury criteria used in the evaluation of performance of automotive vehicles can be established, such vehicles can be designed for improved occupant protection. Both the injury criteria and the crash tests used in the design and in the evaluation of the car passive safety performance are presented and discussed in the following section. After, the use of these criteria and tests in regulations for the design and analysis of automobiles is presented for scenarios of front, side, and rollover crashes.

42.2 MEASURING OCCUPANT PROTECTION

The design of vehicles for improved passive safety requires that, in one hand, there are numerical and experimental tools for the quantification of the occupant injury during a crash and, in the other hand, standard crash scenarios, representative of the most important types of accidents, are clearly defined. The quantification of the potential for human injury in a vehicle crash is defined by using injury criteria for the different body parts. The standard vehicle crash scenarios are defined in regulations by most of the international traffic authorities and consumer groups.

42.2.1 Injury Criteria for Different Body Parts

The injury criteria used in different regulations for the approval of vehicles, or for rating their performance with respect to passive safety, are reviewed here. It should be noted that many other injury criteria have been proposed, for different body parts, their application being restricted to particular user groups. However, due to the actual criticism received by some of the current injury criteria used

in regulation or because the protection of other body segments becomes relatively more important, it is possible that new criteria get approved or that the existing are reviewed to include more updated scientific knowledge [15].

42.2.1.1 Head Injuries

The *Head Injury Criterion* (*HIC*) is probably the best known of the injury criteria used worldwide. This criterion, using the head translation acceleration, is computed by [16]

$$HIC = \max\left[\frac{1}{t_2 - t_1}\int_{t_2}^{t_1} a(t)\,dt\right]^{2.5}(t_2 - t_1) \tag{42.1}$$

$a(t)$ being the head acceleration, $(t_2 - t_1)$ a predefined window on time during the crash event, and *max* the largest value obtained for the integral in any time window during the crash event. Some regulations require that $(t_2 - t_1) = 36$ ms, for which the criterion is designated as HIC_{36}. The maximum value for the HIC_{36}, or for its European equivalent Head Protection Criterion (*HPC*), is 1000 for both Federal Motor Vehicle Safety Standards (FMVSS) and Economic Council for Europe (ECE) regulations. In some regulations the period of time used for the calculation of the criterion is 15 ms, 700 being the maximum for the HIC_{15} when applied to a 50th percentile male.

When dealing with the contact of occupants with the internal structures of the vehicle, the *3 ms criterion* (a_{3ms}) is used in several European and North American regulations. This criterion is defined as the mean acceleration level obtained for an impact duration of 3 ms [5]. The values of a_{3ms} must always be below 80 g. Such as for the *HIC*, also the a_{3ms} is based on the Wayne State Tolerance Curve [17] that only accounts for head translation accelerations.

The attempt to consider translational and angular head accelerations in a single criterion has been proposed by Newman [18] with the *Generalized Acceleration Model for Brain Injury Threshold* (*GAMBIT*), which is expressed as

$$G(t) = \left[\left(\frac{a(t)}{a_c}\right)^n + \left(\frac{\ddot{\alpha}(t)}{\ddot{\alpha}_c}\right)^m\right]^{1/s} \tag{42.2}$$

in which $a(t)$ and $\ddot{\alpha}(t)$ are the head translational and angular accelerations, respectively, n and m are empirical constants to fit the data available, and a_c and $\ddot{\alpha}_c$ are critical values for the translational and angular accelerations. Although this HIC has the potential to include multidirectional accelerations, it has not been validated yet and, consequently, is not included in any of the regulations [19].

42.2.1.2 Spinal Injuries

The injuries on the human spine during traffic accidents have particular incidence for the cervical spine, or neck, and less for the thoracic and lumbar spines. Although the neck injuries appear to be lower than the head injuries in general collisions, for crash scenarios characterized by rear-end contact, a great number of injuries are at the cervical spine level [19]. Even minor injuries may have long-term consequences as, for instance, those due to whiplash.

The *Neck Injury Criterion* (*NIC*), developed by Boström et al. [20], takes into account the changes on fluid pressure on the compartments of the cervical spine and relates those to neck injuries. The *NIC* is defined by

$$NIC(t) = 0.2\,a_{rel}(t) + \left[v_{rel}(t)\right]^2 \tag{42.3}$$

where $a_{rel}(t)$ is the relative acceleration in the anterior–posterior direction, between the head center of mass and the first thoracic vertebra, and $v_{rel}(t)$ the relative velocity. The risk for a minor injury has been found to be limited by 15 m^2/s^2. Although this criterion is used in accidentology studies, its application in many kinematic conditions is questionable [5].

The U.S. National Highway Traffic Safety Administration (NHTSA) proposed the N_{ij} *NIC* to account for the injuries in frontal impacts in severe conditions, as those involving airbag deployment or high accelerations [21], given by

$$N_{ij} = \frac{F_Z}{F_{int}} + \frac{M_y}{M_{int}} \tag{42.4}$$

where F_Z and M_y are the axial and sagittal bending moment, respectively, while F_{int} and M_{int} are the respective critical intercept values, established for appropriate dummies. For instance, F_{int} = 4500 N in compression and tension, M_{int} = 310 Nm in flexion, and M_{int} = 125 Nm in extension are all defined for the 50th percentile dummy [5]. The N_{ij} is calculated for the linear combination of axial forces, tension or compression, and bending moments, extension or flexion. The four values calculated combining an axial force with a moment, designated by N_{te}, N_{tf}, N_{ce}, and N_{cf}, should not exceed 1.0 for each loading combination.

Due to the increasing relative importance of other loading cases and of the prevention of injury for low-speed impacts, other injury criteria, addressing the neck, are being proposed. Among these, the *Neck Protection Criterion* (N_{km}), by Schmitt et al. [22]; the *Intervertebral Neck Injury Criterion* (*IV-NIC*), by Panjabi et al. [23]; the *Neck Displacement Criterion* (*NDC*), by Viano [24]; or the *Lower Neck Load Index* (*LNL*), by Heitplatz [25] address different features of the neck injury mechanisms. However, either due to limited validation, limited range of application, or other reasons, these criteria are yet being investigated and have not been included in any passive safety regulations yet.

42.2.1.3 Thoracic Injuries

After the head and the neck injuries, the thorax is the most critical region to protect in vehicle accidents. Although there are few long-term disabilities due to thorax injuries, the victim either dies soon after or recovers completely [19]. In terms of vehicle passive safety, the important injuries to protect are those resulting from blunt trauma that can result in rib fractures, lung contusion, or vessel disruptions.

The limits on the *acceleration and force* that the thorax can withstand are included in some regulations for frontal impact. The tolerance level for forces on the thorax is 3.3 kN for the sternum and 8.8 kN for chest and shoulders, while the tolerance level for acceleration using the a_{3ms} is 60 g [5]. For lateral impacts the tolerance levels are 7.4 kN for AIS0 and 10.2 for AIS3 [26]. However, Viano [27] pointed out that for a 25% probability of AIS4 or higher, the tolerance level is 5.5 N for forces and 45.2 g for the lateral acceleration on the T8 vertebra or 31.6 g if the lateral acceleration is measured in the T12 vertebra.

For side impact the *Thoracic Trauma Index* (*TTI*) relates the injuries to the maximum lateral acceleration measured in the struck side of the rib cage and lower thoracic spine, taking into account the subject age and weight [5]. The *TTI* is given by

$$TTI = 1.4Age + 0.5(a_{RibY} + a_{T12_Y})\left(\frac{M}{M_{std}}\right) \tag{42.5}$$

where

Age is the age of the subject in years
a_{RibY} is the maximum lateral acceleration of 4th and 8th ribs on the struck side
a_{T12Y} is the maximum absolute acceleration of the 12th thoracic vertebra
M is the subject mass
M_{std} = 75 kg. The threshold values for *TTI*, in some regulations, are 85 g in a SID dummy for a four-door car and 90 g for a two-door car

The thorax *Compression Criterion* (*CC*) has been shown to correlate well with the AIS [5]. The chest deflection is measured for the 50th percentile male, with and original chest size of 230 mm. For instance, a compression of 40% of the chest correlates well with a severe injury AIS4.

It is observed that the soft tissue injury is not only compression dependent but also time–rate dependent. The *Viscous Criterion* (*VC*), which correlates well with the risks for thoracic injuries [28], is given as

$$VC = v(t)\, c(t) \tag{42.6}$$

in which $v(t)$ is the velocity of deformation of the thorax and $c(t)$ is the instantaneous compression function calculated as the ratio of the thorax deformation by the initial torso thickness. The maximum of the *VC*, VC_{max}, correlates well with injury risk. For a 25% probability of severe injuries, AIS4 or larger, the tolerance level is VC_{max} = 1 m/s [28], which is the limit value used in regulations.

By putting together the compression and acceleration responses, the *Combined Thoracic Index* (*CTI*) for frontal impact is defined as [21]

$$CTI = \frac{A_{max}}{A_{int}} + \frac{D_{max}}{D_{int}} \tag{42.7}$$

where

$A_{max} = a_{3ms}$ of the resultant acceleration of the spine in g
A_{int} the critical intercept value which is 85 g for the 50th percentile Hybrid III dummy
D_{max} the deflection of the chest in mm
D_{int} the critical intercept value, which is 102 mm for the same dummy

42.2.1.4 Abdominal Injuries

A large amount of work is being put on the development of suitable criteria for abdominal injuries. The strength and physiological alterations of the internal organs in the face of external loading or accelerations are being tested in order to define thresholds for injury and suitable descriptors [5]. However, owing to the complex nature of the problem, not many conclusions on the use of injury criteria are published for abdominal injuries, the *Abdominal Peak Force* (*APF*) being the only measure used in regulations. The *APF* is determined by using the Euro-SID dummy and it is required to be less than 2.5 kN, internal force, or 4.5 kN, external force.

42.2.1.5 Pelvic Injuries

Clinically the pelvic injuries are characterized by fractures of the pelvic ring and sacrum or associated injuries [5]. It has been shown that forces higher than 10 kN and accelerations in excess of 100 g can be sustained if the loading is done through the greater trochanter and iliac wing [29]. This leads to the *Pelvis Lateral Acceleration (PLA)* criterion, used in the standard FMVSS 214, that specifies a limit of 130 g can be withstood by the pelvis before injury. However, it should be noted that the threshold set by the PLA criterion is for loading of the whole pelvis. If only the greater trochanter is loaded, fracture can occur at about half of this threshold.

42.2.1.6 Lower Extremity Injuries

The large improvements in head, neck, and thorax protection and the consequent decrease in fatalities due to injuries in the upper body made the lower extremity injuries relatively more important in recent years. The *Tibia Compression Force Criterion* (*TCFC*) is defined as the maximum force transmitted to each tibia of the test dummy along its axial direction. The maximum value acceptable for the *TCFC* is 8 kN in some European regulations. The FMVSS regulations specify a maximum compression force of 10 kN for the hip–thigh–knee [5].

The *Femur Force Criterion* (*FFC*) uses the compression force on the femur and its duration, measured in ms. The European regulations specify that the *FFC* must be under 7.58 kN if applied longer than 10 ms.

Other criterion, resulting from static bending testing by Yamada [30], is the *Tibia Index* (*TI*), which is calculated as

$$TI = \frac{M}{M_{crit}} + \frac{F}{F_{crit}} \tag{42.8}$$

in which M is the bending moment and F the compressive force, while M_{crit} and F_{crit} are intercept values that take the values of 225 N m and 33.9 kN, respectively, for the 50th percentile male. The European regulations specify that *TI* must not exceed 1.3 in each tibia.

42.2.1.7 Relation between Injury Criteria and Occupant Protection

The injury criteria scores are not always directly related to the existence of injuries in case of a crash but specify thresholds for occupant protection to which some rating may be associated. A relation between injury criteria scores and protection levels is described in Table 42.3 [31].

At this point it should be noted that the relations between the scores of the injury criteria and the actual mechanisms of injury are not generally well mapped. Even the relations between the scores of the AIS and the injury criteria are not clearly defined, although there is a large body of work done to relate them. The interested reader is referred to the work of Schmitt, Niederer, and Walz [5] for more detailed information on this subject.

TABLE 42.3
Injury Parameter Cutoff Values Associated with Possible Injury Protection Ratings

Body Region	Parameter	IARV[a]	Good–Acceptable	Acceptable–Marginal	Marginal–Poor
Head and neck	HIC_{15}	700	560	700	840
	N_{ij}	1.0	0.8	1.0	1.2
	Neck axial tension (kN)	3.3	2.6	3.3	4.0
	Neck contraction (kN)	4.0	3.2	4.0	4.8
Chest	Thoracic spine acceleration (a_{3ms})	60	60	75	90
	Sternum deflection (mm)	−50	−50	−60	−75
	Sternum deflection rate (mm/s)	−8.2	−6.6	−8.2	−9.8
	VI	1.0	0.8	1.0	1.2
Leg and foot	*FFC*	−9.1	−7.3	−9.1	−10.9
	Tibia–femur displacement (mm)	−15	−12	−15	−18
	TI	1.0	0.8	1.0	1.2
	TCFC	−8.0	−4.0	−6.0	−8.0
	Foot acceleration (g)	150	150	200	260

[a] IARV—Injury assessment reference value.

42.2.2 Regulations for Automotive Crash Testing

Real-world crashes are complex events for which it is not possible to define crash scenarios to foresee all possible combinations of outcomes. Using the background of the accident analysis and the statistical data available, such as that presented in Figure 42.6, it is possible to set the priorities for occupant protection [10,12]. The definition of the limits of human tolerance to impact and the identification of the means to measure the biomechanical outcomes of impact led to the establishment of performance criteria that can be effectively used in vehicle design and analysis. The definition of a number of crash scenarios representative of the most important classes of accidents that can be standardized, enabling the assessment of the vehicle safety by means of biomechanical criteria, was the next necessary step, which led to the passive safety standards used today [4,5,10]. The requirements for some of these standards are summarized in Tables 42.4 through 42.6.

The regulations for passive safety tests, which all vehicle models must pass before being approved, reflect the priorities in occupant protection defined with the mechanisms described before. The official standards in Europe are defined by the ECE and designated by ECE R++, while in the United States of North America, the regulations are defined by the NHTSA and comprised in the FMVSS, designated by FMVSS+++. Also in Asia and Oceania, there are several organizations responsible for issuing regulations, which are mostly similar to their European and North American counterparts. Also the different New Car Assessment Programs (NCAPs) have sets of testing programs that differ, to a certain extent, from those used by the different official organizations. Due to their public acceptance, the EuroNCAP tests are listed here. Notice though that the harmonization of the NCAP tests is not yet done [32].

The frontal impact test regulations for the ECE, FMVSS, and EuroNCAP are summarized in Table 42.4. In all tests the Hybrid III dummy is used and the deformable barrier used in some of the tests is completely defined in the regulation also.

The side impact regulations for the ECE, FMVSS, and EuroNCAP are summarized in Table 42.5. All tests use impacting barriers, with characteristics defined in the standards, and side impact dummies. Notice both dummies and barriers used for the European and for the North American tests are different. In particular, the barriers not only have different geometries and material properties but also have different heights, therefore engaging different parts of the car.

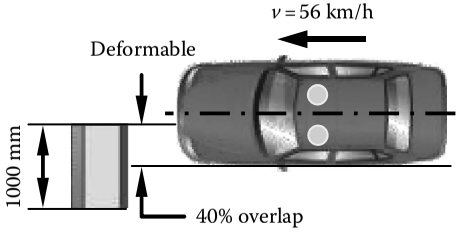
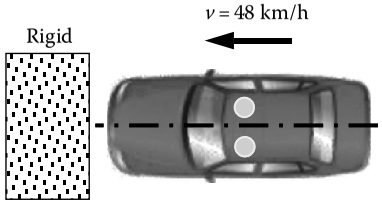
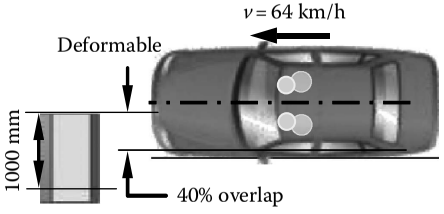
The pole test for side impact is included in the FMVSS regulations but not in the European regulations. The EuroNCAP test also uses the pole impact test as an optional form of evaluating the side protection of vehicles, for all those that achieve a maximum score in the standard's side impact test. These tests, described in Table 42.6, assess effects of the intrusion of the side of the vehicle on the occupant safety and can represent the contact not only with stationary objects but also with moving vehicles, such as motorcycles.

The rear impact tests, used in the European and North American regulations, described in Table 42.7, are designed to address the stability of the fuel system only. However, with the growing relative importance of the occupant injuries resulting from rear-end strikes, such as the whiplash, it is foreseen that they will be introduced in legislation as a form to quantify occupant safety protection also.

Besides the front, side, and rear-end impact tests, the rollover tests are also used in regulation but with the objective of assessing the fuel container performance, roof crush resistance, and windshield retention, as in FMVSS216 and FMVSS208. Several regulations are also in place to define the compliance of the exterior of the vehicle with femur and head impacts of vulnerable road users [41,42].

It should be noted that although the full-scale tests imposed by the regulations are of fundamental importance, they serve as a reliable final check but not as tools for the development of new car models [43]. The standards are translated into a set of component performance ratings that are checked by specially developed tests that make part of internal standards of each car company. Ardoino [43] refers that the development of a car in FIAT required about 70 full-scale crash tests; 80 crash simulation tests, that is, sled tests; 250 component tests; and over 50 virtual tests in a computer. With the

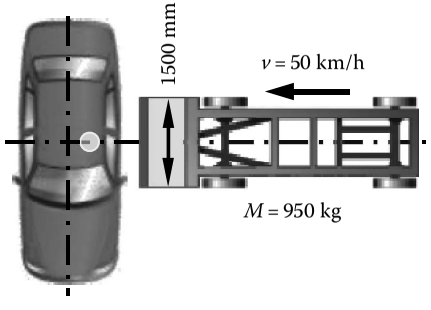
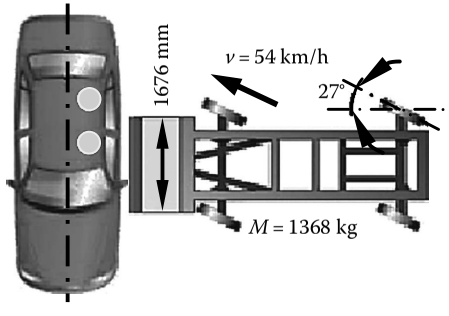
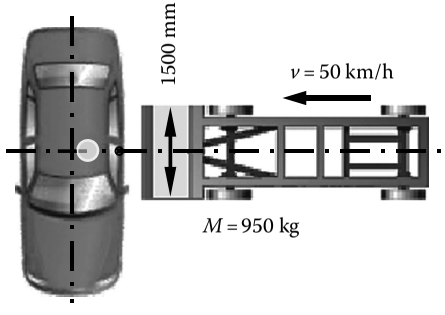
TABLE 42.4
Front Impact Tests in Regulations and Threshold Values for Injury

	European Regulation R94 [33]	U.S. Regulation FMVSS 208 [34]	New Car Assess. Prog. EuroNCAP[a] [35]
Test configuration			
Dummies	2 Hybrid III	2 Hybrid III	2 Hybrid III
Head injury	HPC < 1000 $a_{3ms} < 80$ g	$HIC_{36} < 1000$ $HIC_{15} < 700$	$a_{peak} < 80$ g If $a_{peak} > 80$ g, $a_{3ms} < 72$ g, and $HIC_{36} < 650$
Neck injury[b]	$N_{extension} < 57$ N m $N_{tension@60ms} < 1.1$ kN $N_{shear@45ms} < 1.1$ kN	$N_{ij} < 1.0$	$N_{extension} < 42$ N m $N_{tension@0ms} < 2.7$ kN $N_{shear@0ms} < 1.9$ kN
Thorax injury	$CC < 50$ mm $VC < 1.0$ m/s	$CC < 76.2$ mm $a_{3ms} < 60$ g	$CC < 22$ mm $VC < 0.5$ m/s
Femur injury+	$FFC_{@10ms} < 7.58$ kN	$FFC < 10$ kN	$FFC_{@10ms} < 3.8$ kN
Knee injury	$\Delta d < 15$ mm		$\Delta d < 6$ mm
Tibia injury	$TCFC < 8$ kN $TI < 1.3$		$TCFC < 2$ kN $TI < 0.4$
Steering wheel displacement	$\Delta x < 100$ mm $\Delta y < 80$ mm		$\Delta y < 72$ mm $\Delta x < 100$ mm
Pedal displacement			

[a] The EuroNCAP limits are presented for maximum score. The R94 limits lead to null score.

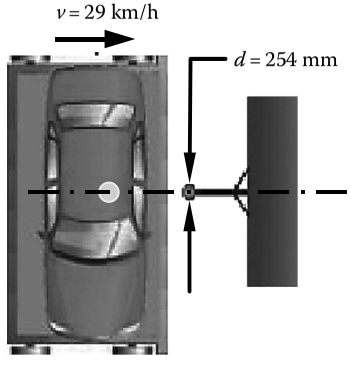
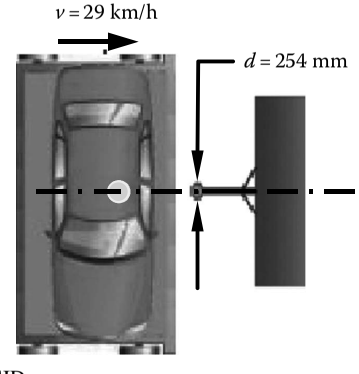
[b] The limits are bounded by a force corridor described in the regulation.

TABLE 42.5
Side Impact Tests in Regulations and Threshold Values for Injury

	European Regulation R95 [36]	U.S. Regulation FMVSS 214 [37]	New Car Assess. Prog. EuroNCAP[a] [35]
Test configuration			
Dummies	1 Euro-SID	2 SID	1 Euro-SID
Head injury	$HPC < 1000$		$a_{peak} < 80$ g If $a_{peak} > 80$ g, $a_{3ms} < 72$ g, and $HIC_{36} < 650$
Thorax injury	$VC < 1.0$ m/s $RDC < 42$ mm	$TTI < 85$ g	$CC < 22$ mm $VC < 0.32$ m/s
Pelvic injury	$PSPF < 6$ kN	$a_{peak} < 130$ g	$PSPF < 3$ kN
Abdomen injury	$APF < 2.5$ kN		$APF < 1.0$ kN m
Tibia injury			$TCFC < 2$ kN $TI < 0.4$

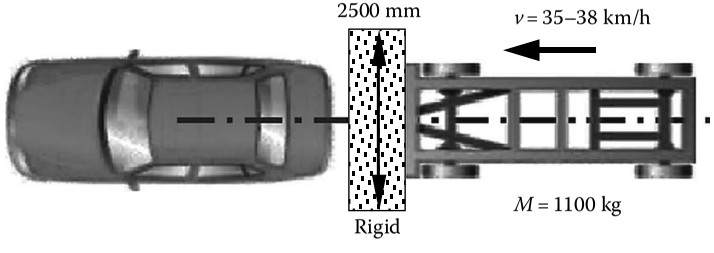
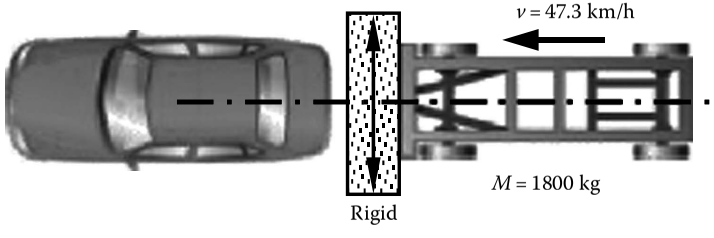
[a] The EuroNCAP limits are presented for maximum score. The R95 limits lead to null score.

TABLE 42.6
Pole Impact Tests in Regulations and Threshold Values for Injury

	European Regulation —	U.S. Regulation FMVSS 201 [38]	New Car Assess. Prog. EuroNCAP[a] [35]
Test configuration	—		
Dummies	—	1 SID	1 Euro-SID
Head injury	—	$HIC_{36} < 1000$	$HIC_{36} < 1000$ $a_{peak} > 80$ g

[a] The EuroNCAP limits are presented for maximum score. The pole test is optional only if side impact test has maximum score.

TABLE 42.7
Rear Impact Tests in Regulations

	European Regulation R32 [39]	U.S. Regulation FMVSS 303 [40]
Test configuration		
Comments	To evaluate fuel container integrity	To evaluate fuel container integrity

evolution of computer technology and fidelity of computer models, the balance between the numbers presented is changing dramatically, but still the idea that component testing, virtual or experimental, is required and not subjected to international regulations is valid.

42.2.3 Anthropometric Testing Devices: Dummies

Regardless of the crash tests being experimental or virtual, the use of Anthropometric Testing Devices (ATDs), better known as crash test dummies, is fundamental to evaluate the vehicle performance in terms of passive safety. The ATD should not be confused as human substitutes but simply as measuring devices as they are intended to measure, with biofidelity, the crash response in terms of accelerations, forces, and deflections of selected points. All dummies have dimensions and masses based on their human counterparts. Each ATD is designed to be used for crash tests in a given direction and cannot, or should not, be used to measure the crash effects in general directions. The development of omnidirectional dummies is still a research topic that is not close to be implemented in any practical application. Certainly the ongoing research on advanced human body models and the understanding of biologic material behavior and physiologic and neurological resistance of the human structures will provide important advancements in such research [44,45].

The frontal impact dummy that is used in all regulations related to frontal impact is the 50th percentile male Hybrid III dummy, shown in Figure 42.7. In various testing setups, the 5th percentile female and the 95th percentile male dummies are also used to cover most of the human population. Also the Test device for Human Occupant Resistant (THOC) dummy is used for frontal impact, although it is not applied in any official regulation.

Contrary to the situation in what front impact dummies are concerned, for side impact the FMVSS and the ECE regulations prescribe the use of different side impact dummies, that is, they prescribe the use of the U.S.-SID and Euro-SID1 dummies presented in Figure 42.8, respectively. In Figure 42.9, numerical models of the side impact dummies, using the finite element (FE) method and multibody (MB) dynamics procedures, are displayed. Some of the major differences between the Euro-SID and the U.S.-SID concern the mechanical components of the torso. Due to the many differences between the dummies, their measuring capabilities, and respective biofidelity, several efforts are being made to obtain unified side impact dummies that can be used worldwide. The evolution of the European dummy (Euro-SID2), the Biofidelic Side Impact Test Dummy (Bio-SID), and the world-SID represents some of the developments in this direction.

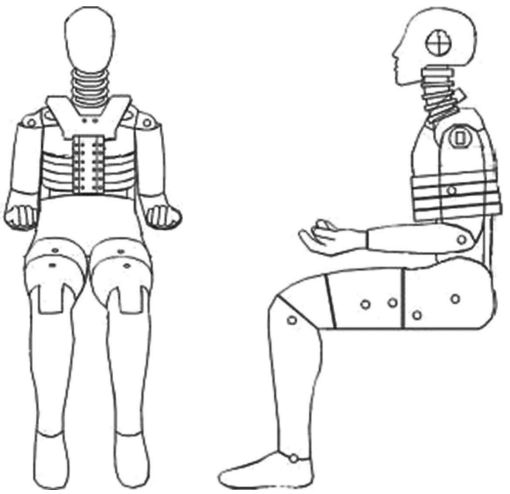


FIGURE 42.7 Characteristic of the 50th percentile Hybrid III dummy. (From National Highway Traffic Safety Administration, Code of Federal Regulations Part §571.208, pp. 497–580, 2004.)

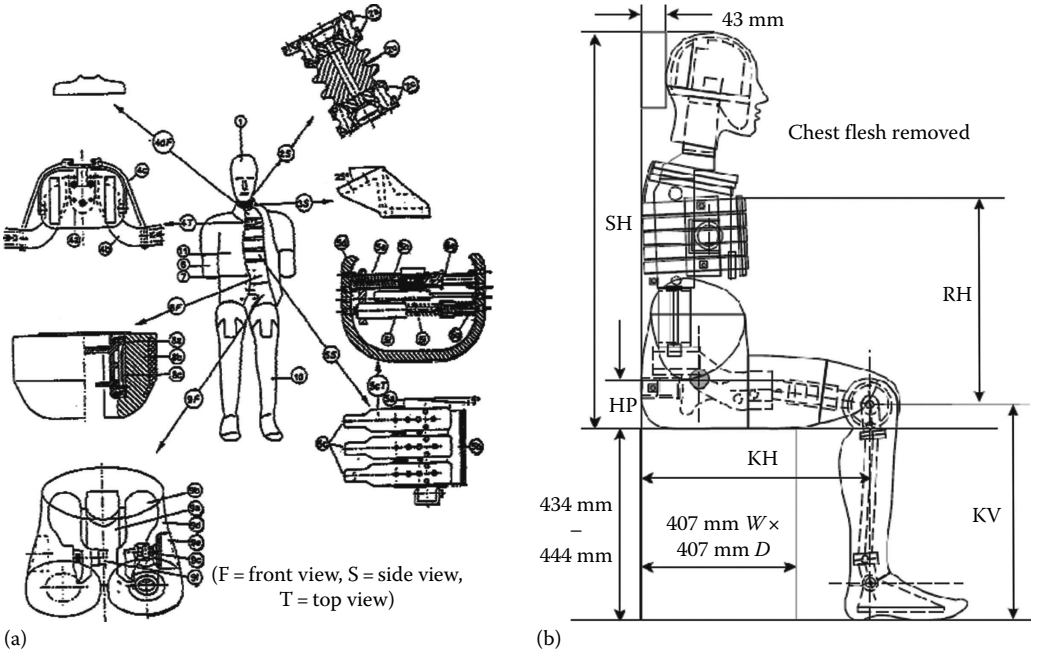


FIGURE 42.8 Characteristic of the side impact dummies: (a) Euro-SID1 (From European Council for Europe, Regulation 95–Uniform provisions concerning the approval of vehicles with regard to the protection of the occupants in the event of a lateral collision, 1995); (b) U.S.-SID. (From National Highway Traffic Safety Administration, Federal motor vehicle safety standard no 214, Side impact protection, Code of Federal Regulations Part §571.214, pp. 666–677, 2004.)

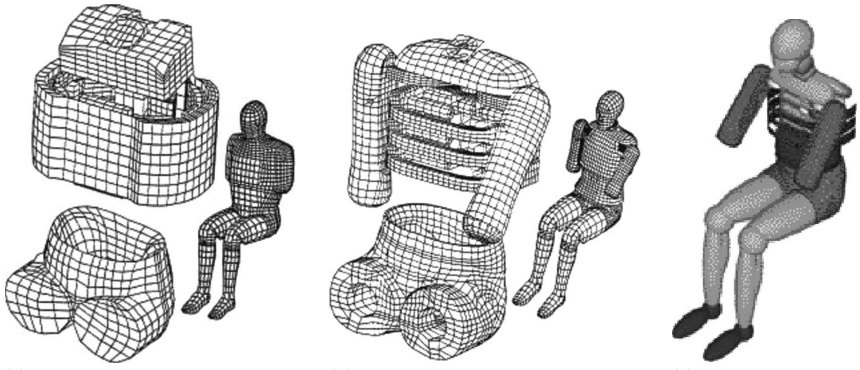


FIGURE 42.9 Numerical models of the side impact dummies: (a) Euro-SID1 (From Lescheticky, T. and Gholami, T., USSID and EuroSID: Two tools for the side impact simulation of car structures, In: J. Ambrósio and M. Pereira eds., *Proceedings of the NATO-ASI on Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection: Volume II—Contributed Papers*, Tróia, Portugal, pp. 19–34, July 7–19, 1996); (b) U.S.-SID (From Lescheticky, T. and Gholami, T., USSID and EuroSID: Two tools for the side impact simulation of car structures, In: J. Ambrósio and M. Pereira eds., *Proceedings of the NATO-ASI on Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection: Volume II—Contributed Papers*, Tróia, Portugal, pp. 19–34, July 7–19, 1996); and (c) MB of the Euro-SID. (From Wismans, J., Design tools: Human body modelling, In: J. Ambrósio, ed., *Crashworthiness: Energy Management and Occupant Protection*, Springer, Vienna, Austria, pp. 333–377, 2001.)

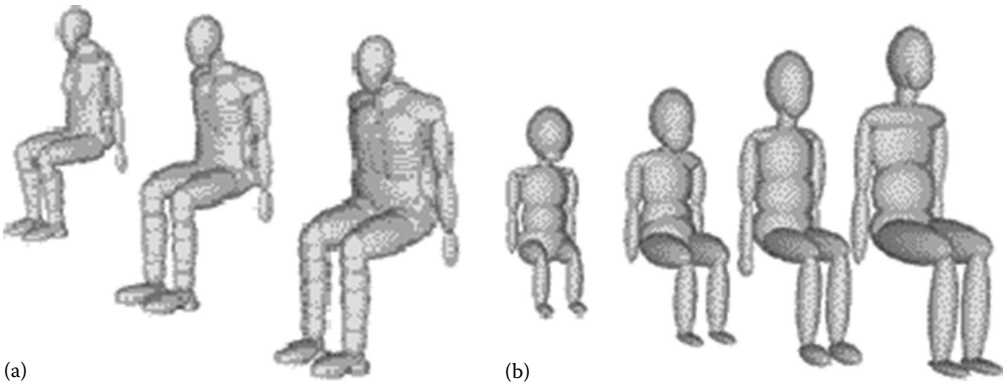


FIGURE 42.10 Examples of MB crash dummy models: (a) Hybrid III dummy family (top), 5th%, 50th%, and 95th%; and (b) CRABI 12-month and TNO child dummies (bottom), P3/4, P3, P6, and P10. (From Wismans, J., Design tools: Human body modelling, In: J. Ambrósio, ed., *Crashworthiness: Energy Management and Occupant Protection*, Springer, Vienna, Austria, pp. 333–377, 2001.)

Although only some of the dummies are used in official regulations, there are complete families developed and used commercially, such as those presented in Figure 42.10. Besides the physical ATD sold by different commercial companies, the software developers also include fully validated models of these families of dummies together with their software. Basically all FE and MB commercial codes used for passive safety include these dummy models [48–51].

Besides the front and side impact dummies, there are dummies also developed for rear-end impact and for other interior special purpose tests. Although the Hybrid II has been used for rear-end impact, its neck is not biofidelic enough for this crash scenario. The Biofidelic Rear-End Dummy (BioRID) has an improved spinal representation that presents biofidelic kinematics for rear-end impacts [5]. Several efforts are also being made to develop suitable standing pedestrian and passenger dummies. The POLAR dummy is an example of these attempts. However, not only is there still a large amount of research to be done before these dummies provide reliable and biofidelic responses, but also regulations requiring their use have not yet been released.

42.2.4 Impact Barriers

The impact barriers play a fundamental role in the crash tests in regulations as normalized representations of collision partners. Both the FMVSS and the ECE regulations prescribe the complete geometry, mass, and material characteristics of the impact barriers. For instance, the regulation FMVSS214 for side impact protection specifies the impact barrier shown in Figure 42.11.

The ECE R94, for frontal impact, specifies not only the geometric and material characteristics for the impact barrier but also the specifications for its construction. The certification of the barrier includes collecting samples at prescribed locations and ensuring that their force–deformation performance matches the specifications listed in the standard [33]. The procedure described in the standard must be followed for physical construction of the impact barrier and a proper equivalent mathematical model must be used to validate any virtual deformable barrier that is used in numerical models. Similar procedures, properly specified in the respective regulations, should be followed for all other types of impact barriers used in crash tests.

42.3 STRUCTURAL FEATURES AND SYSTEMS FOR PASSIVE SAFETY

The passive safety of vehicle occupants starts with the design of vehicle structures, able to exhibit controlled deformations with suitable energy absorption characteristics and undeformed, or with limited intrusion, occupant survivable cells. After ensuring the proper characteristics of the vehicle structure,

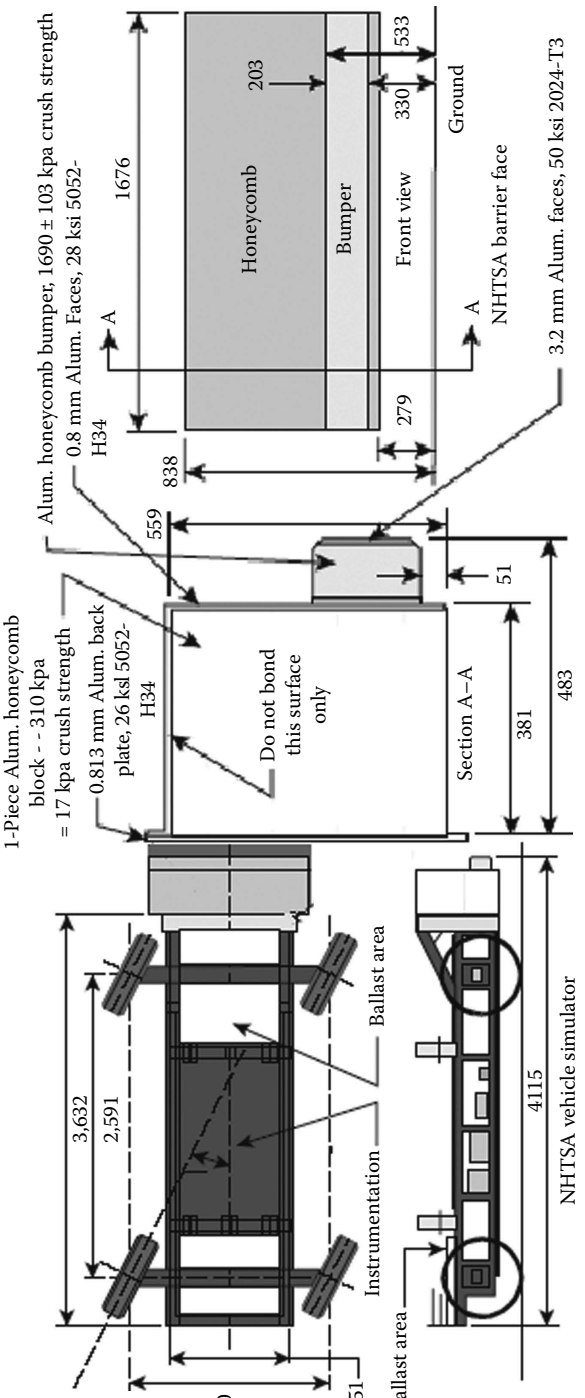


FIGURE 42.11 Impact barrier for the FMVSS214 regulation. (From National Highway Traffic Safety Administration, Federal motor vehicle safety standard no 214, Side Impact Protection. Code of Federal Regulations Part §571.214, pp. 666–677, 2004.)

generally in the body in white, safety equipment, such as seat belts and airbags, can be installed to ensure an effective protection of the occupant. The structural requirements for the vehicles protection in frontal and side impacts and for rollover are overviewed here. A brief overview of protection systems, emphasizing seat belts and airbags, is presented afterward.

42.3.1 Vehicle Structure for Crashworthiness

Although the body of the vehicle is primarily designed to carry the powertrain, occupants, and cargo, its passive safety performance has become more and more important. Regardless of the many variations on the arrangements of the front structure of a vehicle, observed in cars by different manufacturers, the main structural components responsible for the controlled deformation and energy absorption are the longitudinal beams, shown in Figure 42.12a.

In a front crash, other structural parts of the vehicle perform a role in the energy absorption characteristics of the vehicle. A typical distribution of the energy absorption characteristics of the front structural components is presented in Figure 42.12b [52]. According to Wittemann [52], the longitudinal beams are responsible for about 50% of the total energy absorption capability of the front of the vehicle, the remaining 50% being the energy absorbed by the deformation of the bumper, front panel, wheel wells, engine, etc. Moreover, typically 50% of the energy absorption occurs in the first half of the vehicle front, that is, between the bumper and the engine, while the remaining 50% of energy are absorbed in the second half of the front, that is, between the engine and the firewall.

Most frontal crashes are either offset or oblique crashes, as shown in Figure 42.6. Regardless of the structural arrangement of the vehicle front that ensures that, even when the crash is offset, the complete front energy absorbing structure is still loaded, the relative contribution of the different vehicle front structural components to the total energy dissipation varies with impact orientation and with the vehicle velocity. Table 42.8 presents a typical distribution of the energy absorption by the different structural elements in frontal crashes [53].

Note that in frontal crashes the survivable space for the occupants must be maintained and the windshield preserved. Furthermore, at least one of the doors must open without tools after the crash. To achieve the integrity of the structure and systems to the rear of the firewall and A-pillar, the sidewalls must be connected to the wheel housing and the force applied to the A-pillar must be transmitted by the doors and side beams to the B-pillar [10].

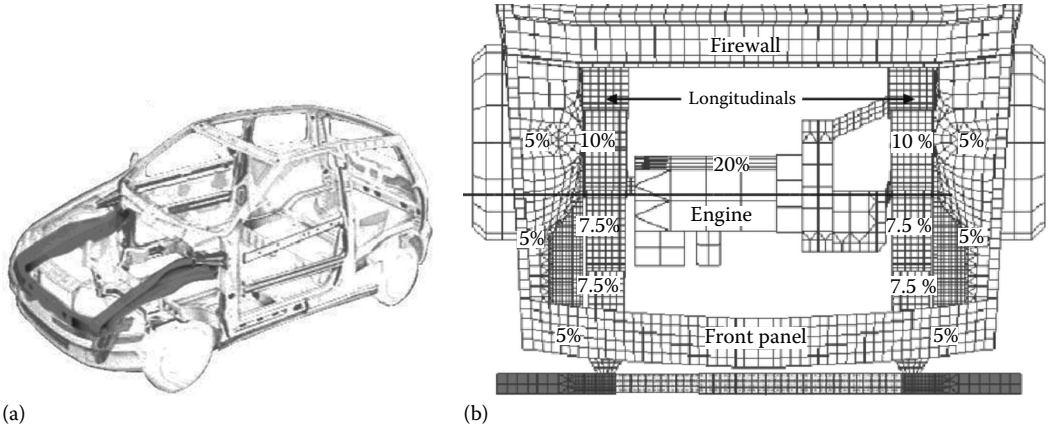
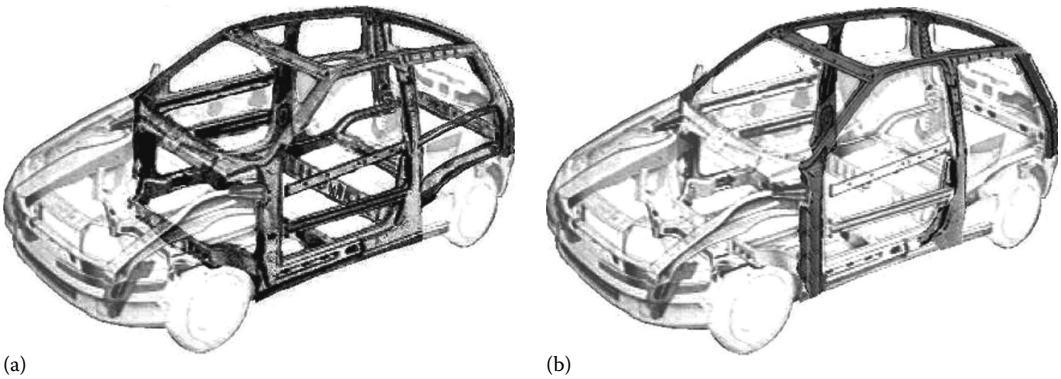


FIGURE 42.12 Energy absorption components in the front of the vehicle for frontal impacts: (a) crash-relevant structural reinforcements in a typical small car; (b) typical energy absorption characteristics of the structural components of the vehicle front. (From Wittemann, W.J., Improved vehicle crashworthiness design by control of the energy absorption for different collision situations, PhD dissertation, Technical University of Eindhoven, Eindhoven, the Netherlands, 1999.)

TABLE 42.8
Energy Absorption Characteristics of the Vehicle Front Structural Components

Structural Component	Full Frontal Barrier		40% Offset Frontal Barrier	
	32 km/h (%)	56 km/h (%)	32 km/h (%)	56 km/h (%)
Bumper	29	24	29	26
Hood	3	4	7	8
Radiator	1	3	3	5
Radiator suspensions	2	2	4	4
Fenders	1	2	2	3
Wheel wells	16	16	15	18
Longitudinal rails	44	45	36	28
Engine	3	3	3	9

Source: Wågström, L. et al., *Int. J. Crashworthiness*, 10(4), 371, 2005.



(a) (b)

FIGURE 42.13 Energy absorption components in the vehicle for side and roof impacts: (a) structural reinforcements for side impacts; (b) structural reinforcements for rollover protection.

Historically, after the front protection, the side impact became the highest priority in vehicle passive safety. The typical structural components of the vehicle responsible for limiting intrusion to the occupant compartment and managing the structural energy absorption are shown in Figure 42.13a. It should be noted that the complete elimination of intrusions in the survivable space of the occupants is not possible for general side impacts, but it is clearly excessive for pole impacts. The strength characteristics of the A- and B-pillars, door reinforcements, strong hinges, efficient locks, and good connections of the doors to sill and rocker of the chassis are some of the structural characteristics of major importance in side impact.

The importance of rollover fatalities in vehicle accidents is much lower than what has been observed for front and side impact but is becoming more and more important as the other passive safety problems are being solved. The typical structural components of the vehicle responsible for the occupant protection in a rollover are shown in Figure 42.13b. The actual regulations only impose the static strength of the roof to be verified, by the regulation FMVSS216, and prescribe a dynamic rollover test to ensure the windshield integrity and that the fuel tank and lines have no leakages.

42.3.2 Vehicle Interior Protection Systems

The most important protection safety systems in the interior of the vehicle are the seat belts and the airbags, followed by the seats and headrests. Although technological developments in the design of more efficient and intelligent occupant protection systems are occurring at a very fast pace, their objectives remain the same, that is, to guide the occupant kinematics to a more favorable configuration in order to mitigate injury. To achieve this objective, both the restraint systems and the airbags require crash sensors, crash sensing criteria, inflators and bags for the airbags, pretensioners, load limiters and belts for the seat belts, and a suitable vehicle interior [54]. In this text only some of the principles of the functions of the protection systems in the vehicle interiors are reviewed. For more detailed information, the reader is referred to the works of Chan [54] on the fundamentals of crash sensing and Seiffert and Wech [10] that addresses general occupant protection. Virtually all commercial vehicles sold today use three-point seat belts for the front occupants in a configuration shown in Figure 42.14a. The anchorage points of the seat belts are generally located in the B- and C-pillars, for the front and rear seat belts, respectively, as shown in Figure 42.14c. In some arrangements, the anchorage points are placed in the seats.

The seat belts with retractors allow controlling the kinematics of the occupant, as shown in Figure 42.14c [4]. It can be noticed that although the body of the occupant is well protected, the head may hit the steering wheel. Therefore, the use of airbags is still required together with seat belts, with their size and pressure being larger in the United States than in Europe to protect both belted and unbelted drivers. It is required that the vehicle structure, where the anchorage points are located, is able to withstand the loads due to the operation of the seat belts. Furthermore, in order to prevent excessive loading of the occupants, due to the retraction systems, the seat belts may be equipped with force limiters also [54,55]. Finally, the correct operation of the seat belts requires that the seats, themselves, have their seat plate reinforced to avoid submarining, that is, to avoid that during deceleration the body of the occupant sinks in the seat and the lower seat belt slips to the abdomen of the occupant increasing the risk for internal injuries [5,56,57].

Airbags for driver and front occupant, such as those presented in Figure 42.15a, are mandatory in all new vehicles in the market. The FMVSS208 specifies that the airbags should protect belted and unbelted passengers, while the ECE specifies the performance of the airbag for belted occupants only. Consequently, the airbags fitting vehicles sold in the North American market are larger and operate with higher pressures, as they may be contacted by head and torso, than those fitting vehicles for the European market that are intended only for head contact [5,10].

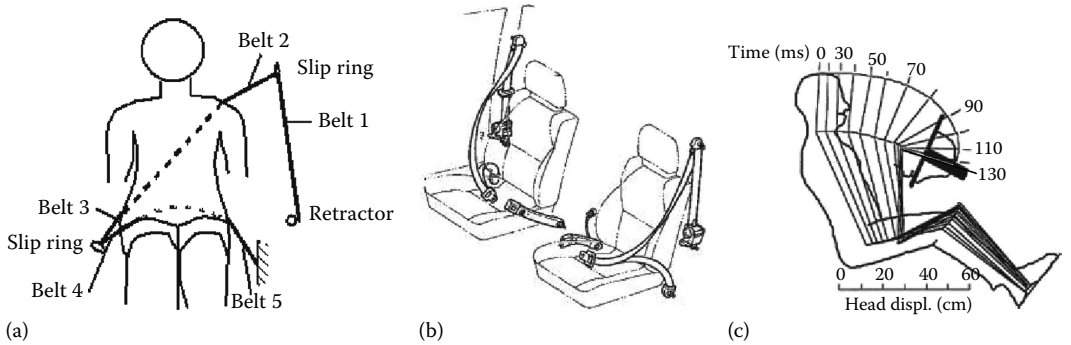


FIGURE 42.14 Seat belts in current vehicles: (a) assembly for the front seats; (b) three-point seat belt with anchorage points (From Introduction in injury biomechanics, In: J. Ambrósio, ed., *Crashworthiness: Energy Management and Occupant Protection*, Springer, Vienna, Austria, p. 305, 2001); (c) occupant excursion with standard retractor seat belt for a 50 km/h sled test. (From Mackay, M., A review of the biomechanics of impacts in road accidents, In: J. Ambrósio, M. Pereira, and F. Pina da Silva eds., *Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection*, Kluwer Academic Publishers, Dordrecht, the Netherlands, p. 115, 1997.)



(a) (b)

FIGURE 42.15 Airbags in a vehicle: (a) front airbags; (b) side airbags for the head and driver thorax.

In lateral collisions, the three-point seat belts do not play a major role in controlling the occupant kinematics. The occupant protection relies mostly in limiting the intrusion of the occupant compartment, on the padding of that compartment, and, if available, on the side airbags, such as the ones presented in Figure 42.15b. The curtain airbags are intended to minimize head injury, while the thorax side airbags provide cushioning for the thorax and lower torso favoring a more compliant contact between the intruding side structure of the vehicle and the occupant. In any case, for both front and side airbags, the airbags need to be contained by the surrounding structure, windows, and windshield.

Besides the seat belts and the airbags, the seats also play a fundamental role on the occupant interior passive safety [5,10]. The seat plate must be reinforced to avoid occupant submarining [56,57]. The geometry of the seat must be such that, in case of side impact, the padding helps support the occupant and protects the pelvic area. Also the headrest plays an important role, especially in rear-end collisions or in the rebound phase of the front crashes, thus helping to minimize whiplash-related injuries. The use of suitable foams for the headrest, able to perform well at different temperatures, and the reduction of the distance between the head and the headrest lead to improvements on the prevention of the whiplash injury [5,58]. Finally, active systems acting over the kinematics of the seats, and able to dissipate some of the kinetic energy of the occupant and seat, are used as optional protection systems in some vehicles [55,59,60].

42.4 NUMERICAL METHODS IN VEHICLE PASSIVE SAFETY

Up until the early 1970, crash studies relied almost exclusively in experimental testing that focused mostly in full-scale testing and in the development of relevant testing scenarios [61]. In a review work, Tidbury [62] stressed the facts that not only the costs associated with experimental testing are very high but also the results cannot be generalized to other impact situations, neither can such methods be used during the design of new vehicles. Therefore, the need for accurate numerical procedures became apparent at that time, and it is even more evident today with the advances in computer power and computational mechanics. The earlier numerical methods used for vehicle crashworthiness were based on the use of lumped masses and nonlinear springs. The models built with these methods, known as lumped parameter models, used lumped masses to represent parts of the vehicle, such as the engine block or the passenger compartment, considered rigid during the analysis, and springs to represent the structural elements responsible for the deformation energy management [63]. The lumped parameter models are mostly 1D or 2D and they include the most very simplified representations of the suspension elements of the vehicles.

The first numerical model for the simulation of the vehicle occupant biomechanics proposed by McHenry [64] was a 2D model that included four articulated rigid bodies and the occupant restraint system. Later, Lobdell [65] proposed a lumped parameter approach with similar numerical characteristics to that used for vehicle structural impact for the simulation of human thorax loading.

The developments of lumped parameter biomechanical models continued through the years of 1970 and 1980 in parallel with the development of models for the structural impact of vehicles but without the synergies associated to the use of a common numerical framework.

The generalization of the vehicle structural and occupant biomechanical models to become fully 3D required another type of numerical methodologies. The MB dynamics methods provided the framework in which more advanced developments could be undertaken. The plastic hinge approach, proposed by Nikravesh and Chung [66] to represent the structural dynamics of road vehicles in crash situations, uses a general MB representation where the full vehicle is described and the structural systems are divided into several rigid bodies connected by kinematic joints to which nonlinear springs, representing the structural deformation, are associated. This approach has been further developed by several authors [67,68] and applied to the crashworthiness of road and rail vehicles proving to be a versatile and efficient methodology to deal with structural crashworthiness, especially in the initial phases of the vehicle design, when extensive reanalysis is required. Still with little or no contact to the developments in structural crashworthiness, the MB methodologies began to be extensively applied in the development of biomechanical models of vehicle occupants. Starting in the initial work of McHenry [64] and going through the developments of Robbins et al. [69] and Fleck et al. [70], the use of MB dynamics methodologies for occupant biomechanics reached a level of maturity that allowed it to become the most popular numerical approach used today in impact biomechanics. The acceptance of the MB-based models by the vehicle industry led to the successful development of commercial occupant simulation computer codes, such as SOMLA [71] and MADYMO[48].

The continuous increase in computer power and the development of the FE method ensured the feasibility of FE-based models for the study of the structural crashworthiness of vehicles. One of the first successful attempts to apply nonlinear FE to the prediction of a vehicle crush was reported by Thompson [72]. Belytschko and Hsieh [73] proposed the FE developments that defined the framework of their use in structural crashworthiness. During the 1970 decade, several nonlinear FE codes suitable to crashworthiness applications were released. Among these the codes DYNA3D by Hallquist [74] and WHAMS by Belytschko [75] are of special importance as they served as the backbone of the codes LS-DYNA [49] and of PAM-CRASH [50] and RADIOSS [51], respectively, which are the most common FE codes used in crash analysis. The application possibilities of this method are being expanded in order to allow for the complex modeling of vehicle and integrated occupants, but still, the representation of the suspension systems of the vehicles and of systems responsible for their active safety is not yet present in the FE models, unless at the cost of a large computational effort. The amount of information required for large systems is still too high, making the application of FEM codes suitable only for the analysis phase of structural and biomechanical systems. For the early design phases, the MB models are still the primary numerical tool used in simulation [67].

Though the features of the different numerical methodologies based on MB dynamics and on FE include most of the ingredients required for the crash analysis, the models more commonly developed neither integrate the structural crashworthiness with the occupant biomechanics nor take into account the overall dynamics of the vehicle. Applications that include either low and medium vehicle impact speeds or multiple impacts or long contact periods or even for which the alignment between vehicles is important for the outcome of the crash are examples of cases for which the vehicle dynamics associated to the suspensions and to the wheel road contact can play an important role. Traditionally the most common numerical techniques for vehicle dynamics modeling are based on MB dynamics methodologies, which are the same for vehicle impact models and for occupant biomechanics modeling. No reason exists for not taking advantage of their features to develop accurate and efficient vehicle models for passive safety.

In this work it is shown that procedures based on MB dynamics provide a unified framework for the development of design and analysis tools of crashworthy structures and of occupants. The models developed are computationally efficient in today's computers to allow for the extensive

reanalysis cycles required in the optimal design of vehicles for crashworthiness. Furthermore, in applications for which the structural patterns of deformation are not known beforehand or for which subcomponent FE models are available, the MB dynamics methods enable the effective coupling between the lumped parameter models and the FE description of the structural deformations.

42.5 MULTIBODY DYNAMICS-BASED ANALYSIS TOOLS: BASICS

A typical MB model is defined as a collection of rigid or flexible bodies that have their relative motion constrained by kinematic joints and that are acted upon by external forces, as shown in Figure 42.16. In the particular case of a vehicle, many MB subsystems can be identified corresponding to suspensions, engine, transmission, body components, and interior components.

The generic MB system is made of *nb* bodies. For the sake of simplicity, in what follows all bodies in the system are considered rigid bodies. The equations of motion for the system of unconstrained rigid bodies are represented by [77]

$$\mathbf{M}\ddot{\mathbf{q}} = \mathbf{g} \tag{42.9}$$

where

M is the mass matrix, which includes the masses and inertia tensors of the individual bodies
$\ddot{\mathbf{q}}$ is the acceleration vector
g the vector with applied forces and gyroscopic terms

The rigid bodies describe components of the vehicle for which the deformation can be neglected in the framework of crashworthiness applications.

The relative motions between the bodies are constrained by kinematic joints, which can be mathematically described by a set of *nc* algebraic equations, written in compact form as

$$\Phi(\mathbf{q}, t) = \mathbf{0} \tag{42.10}$$

The kinematic joints are used to represent the restraints that restrict the relative motion between the system components. Hinges on the suspension systems and on the steering system are examples of features that can be modeled with kinematic joints. However, if the deformations on these elements are of importance for the application envisaged, such as the bushings of the suspension joints in stability and maneuverability studies, the kinematic joints are modeled as force elements [77].

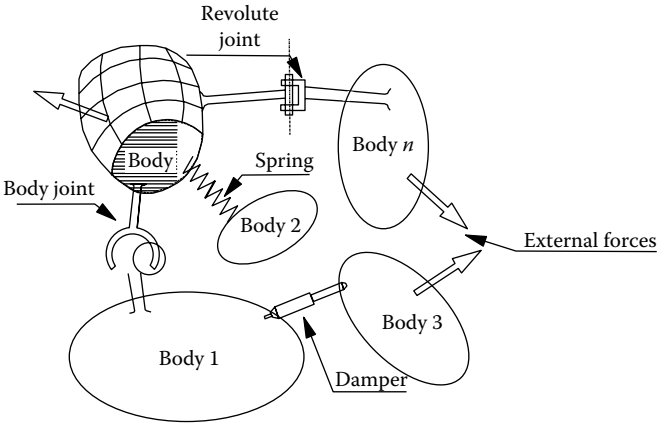


FIGURE 42.16 Generic MB system.

The kinematic joints are also used to represent the deformation mechanisms of structures as part of the plastic hinge technique.

The first and second time derivatives of Equation 1.10 constitute the velocity and acceleration constraint equations, written as

$$\begin{aligned} \dot{\Phi}(\mathbf{q},t) &\equiv \mathbf{D}\,\dot{\mathbf{q}} = \mathbf{v} \\ \ddot{\Phi}(\mathbf{q},\dot{\mathbf{q}},t) &\equiv \mathbf{D}\,\ddot{\mathbf{q}} = \ddot{\gamma} \end{aligned} \tag{42.11}$$

D being the Jacobian matrix. Usually the effect of the kinematic joints for a system of constrained bodies is included in the set of equations (42.9) by adding to the right-hand side the equivalent joint reaction forces $\mathbf{g}^{(c)} = -\mathbf{D}^T\lambda$, where λ is a vector with *nc* unknown Lagrange multipliers, leading to

$$\mathbf{M}\,\ddot{\mathbf{q}} = \mathbf{g} - \mathbf{D}^T\boldsymbol{\lambda} \tag{42.12}$$

Equation 42.12 has *nb* + *nc* unknowns that must be solved together with the second time derivative of the constraint equations. The resulting system of differential algebraic equations is

$$\begin{bmatrix} \mathbf{M} & \mathbf{D}^T \\ \mathbf{D} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{q}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{g} \\ \boldsymbol{\gamma} \end{bmatrix} \tag{42.13}$$

Notice that the solution of Equation 42.13 presents numerical difficulties resulting from the need to ensure that the kinematic constraints are not violated during the integration process.

For the forward dynamic analysis of an MB system, it is required that the initial conditions of the system are given. With $\mathbf{q}^0$ being the position vector and $\dot{\mathbf{q}}^0$ the velocity vector, Equation 42.13 is assembled and solved for the unknown accelerations, which are in turn integrated in time together with the velocities. The numerical methods used must include either constraint violation corrections or their stabilization, especially for long simulation times.

The force vector, appearing in Equation 42.13, includes all forces applied to the components of the vehicle and its occupants described as rigid bodies. Contact forces due to impact, tire/road interaction, springs and dampers, or passive safety restraint systems represent some of the relations that must be included in the force vector. Note that some of the most important modeling efforts of vehicle systems are often done on finding suitable parameters to use the force constitutive relations of the different elements. The development of tire models suitable for different types of automotive studies is just an example of this complexity.

42.5.1 Contact Detection and Contact Models

A set of contact surfaces is defined for the calculation of the external forces exerted on the model when the bodies contact other objects or different body segments. These surfaces are ellipsoids and cylinders with the form depicted by Figure 42.17a for a human body model and in Figure 42.17c for the components of the U.S.-SID crash dummy. When contact between components of the biomechanical model is detected, a contact force is applied to such components in the point of contact. Friction forces are also applied to the contact surfaces using Coulomb friction. Note that the characterization of the surfaces in contact is important for general applications of the biomechanical model.

The numerical procedures for contact detection in crash applications are similar to methodologies used to detect contact in other types of applications. Let a body of the system get close to a surface during the motion of the MB system, as represented in Figure 42.18. Let the impacting surface be described by a mesh of patches. In particular, let the triangular patch, where node *k* of the body

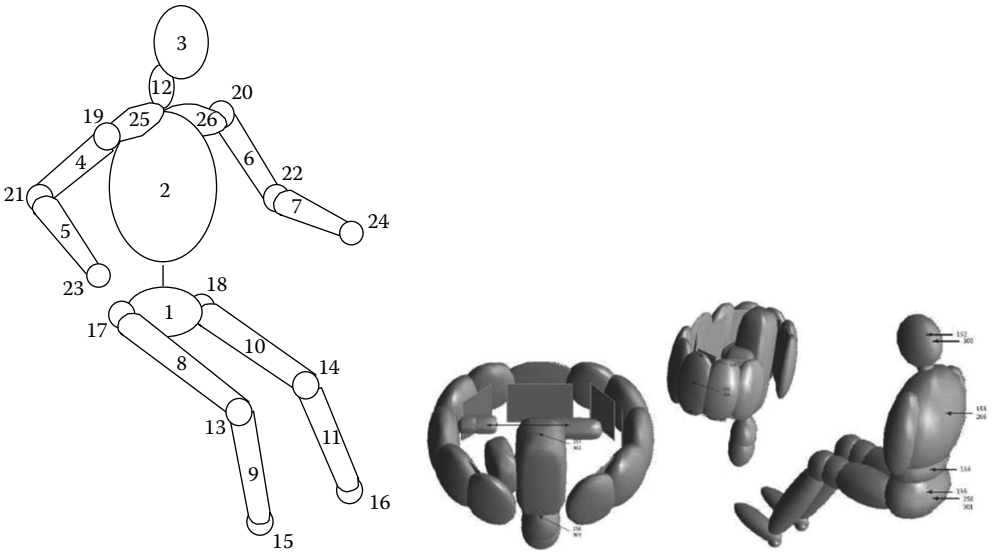


FIGURE 42.17 Representation of contact surfaces in human surrogates: (a) biomechanical model of the human body; (b) U.S.-SID model with internal parts of the thorax and complete model. (From TNO Automotive: MADYMO Manuals, Versions 6.2. TNO MADYMO BV., Delft, the Netherlands, 2004.)

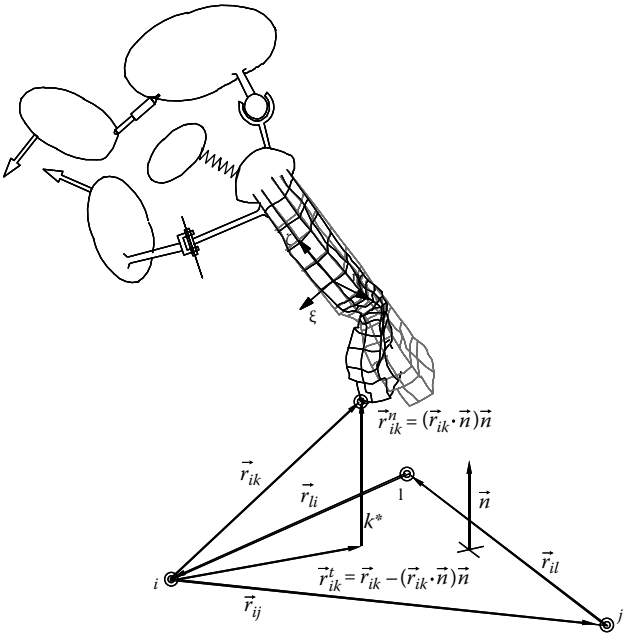


FIGURE 42.18 Detection of contact between an MB component and a triangular patch.

will impact, be defined by points *i*, *j*, and *l*. Note that node in this context means either a point of a rigid body or a nodal point of the FE mesh of a flexible body. If the node (or point) *k* is inside the patch and behind its surface with a penetration $\delta > 0$, there is a contact force applied in the body in its position. Although there are several contact models available to find such forces, a continuous force model is used here.

Let the contact force between two bodies or a system component and an external object be a function of the pseudo-penetration δ and pseudo-velocity of penetration $\dot{\delta}$ given by

$$\mathbf{f}_{s,i} = \left(K\delta^n + D\dot{\delta}\right)\mathbf{u} \tag{42.14}$$

where

K is the equivalent stiffness
D is a damping coefficient
$\mathbf{u}$ is, in this context, a unit vector normal to the impacting surfaces

The hysteresis dissipation is introduced in Equation 42.14 by $D\ \dot{\delta}$, the damping coefficient being written as

$$D = \frac{3K(1-e^2)}{4\dot{\delta}^{(-)}}\delta^n \tag{42.15}$$

This coefficient is a function of the impact velocity $\dot{\delta}^{(-)}$, stiffness of the contacting surfaces, and restitution coefficient e. The generalized stiffness coefficient K depends on the geometry material of the surfaces in contact, which for the contact of a sphere and a flat surface is [78]

$$K = 0.424\sqrt{r}\left(\frac{1-\nu_i^2}{\pi E_i} + \frac{1-\nu_j^2}{\pi E_j}\right)^{-1} \tag{42.16}$$

where ν_l and E_l are Poisson's ratio and Young's modulus associated to each surface and r is the radius of the impacting sphere. The nonlinear contact force is obtained substituting Equation 42.15 into Equation 42.14, written as

$$\mathbf{f}_{s,i} = K\ \delta^n\left[1 + \frac{3\left(1-e^2\right)}{4}\frac{\dot{\delta}}{\dot{\delta}^{(-)}}\right]\mathbf{u} \tag{42.17}$$

This equation is valid for impact conditions, in which the contacting velocities are much lower than the propagation speed of elastic waves, that is, $\dot{\delta}^{(-)} \le 10^{-5}\sqrt{E/\rho}$.

The contact forces between the node and the surface include friction forces modeled using the Coulomb friction model. The dynamic friction forces in the presence of sliding are

$$\mathbf{f}^{friction} = -\mu_d f_d\left(\frac{\left|\mathbf{f}_{s,i}^n\right|}{\left|\dot{\mathbf{q}}_k\right|}\right)\dot{\mathbf{q}}_k \tag{42.18}$$

where

μ_d is the dynamic friction coefficient
$\dot{\mathbf{q}}_k$ is the velocity of point k

The dynamic correction coefficient f_d is expressed as

$$f_d = \begin{cases} 0 & \text{if } \left|\dot{\mathbf{q}}_k\right| \le v_0 \\ \left(\left|\dot{\mathbf{q}}_k\right| - v_0\right)\Big/\left(v_1 - v_0\right) & \text{if } v_0 \le \left|\dot{\mathbf{q}}_k\right| \le v_1 \\ 1 & \text{if } \left|\dot{\mathbf{q}}_k\right| \ge v_1 \end{cases} \tag{42.19}$$

The dynamic correction factor prevents the friction force to change direction for almost null values of the nodal tangential velocity, which would be perceived by the integration algorithm as a high-frequency response, forcing it to reduce the time step size. The friction model represented by Equation 42.19 does not account for the adherence between the node and the contact surface. The interested reader is referred to the work of Wu et al. [79] for a comprehensive discussion on the topics of stiction and sliding in MB dynamics.

42.5.2 Plastic Hinge Approach to Structural Deformations

The methodology reported here uses the plastic hinge approach, suitable for overloaded members in which localized areas of the structure experience the plastic deformations. Plastic hinges usually occur near member's joints, weak areas of the element, and load application points. This observation is important since it allows identifying their location well in advance of the simulation. The concept of plastic hinge is visualized in Figure 42.19 for the bending of a simple supported beam.

At the plastic hinge point, it is necessary to apply a kinematic joint, which describes the kinematics of the deformation, and a generalized spring element that is used to represent the constitutive characteristics of deformation. Such constitutive spring element must represent the elastic–plastic stiffness of the member and the energy absorption characteristics when plastic deformations occur. This spring element can be used with the common types of joints applied, in MB modeling, as depicted in Figure 42.20, for one-axis bending, two-axis bending, and torsion and axial crushing.

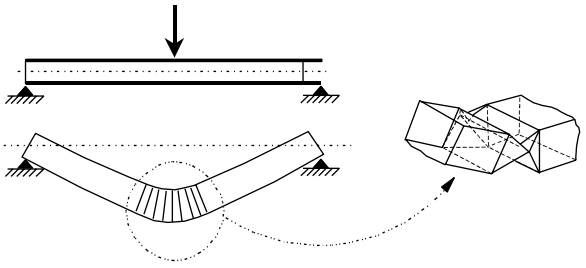


FIGURE 42.19 Localized deformations on a beam and a plastic hinge.

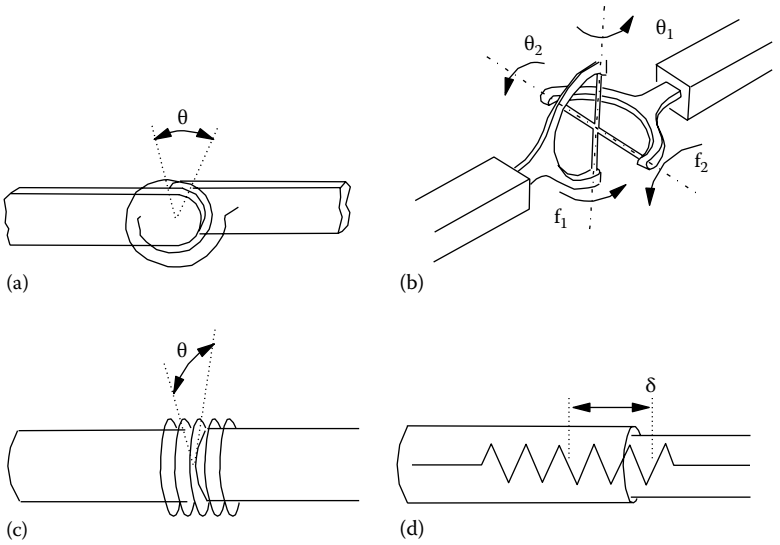


FIGURE 42.20 Plastic hinge models for different load joints: (a) one-axis bending; (b) two-axis bending; (c) torsion; (d) crushing deformation.

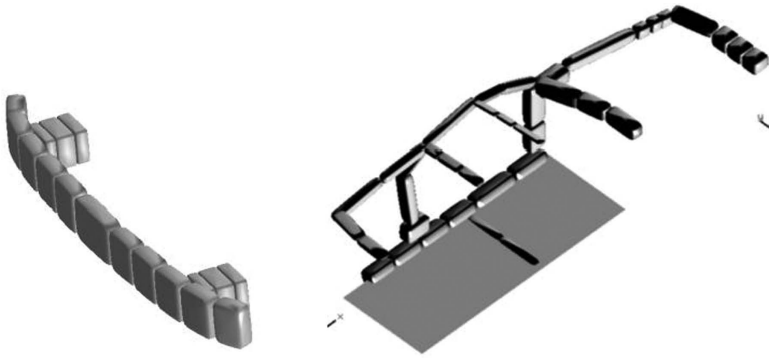


FIGURE 42.21 Plastic hinge models for vehicle substructures: bumper and frame.

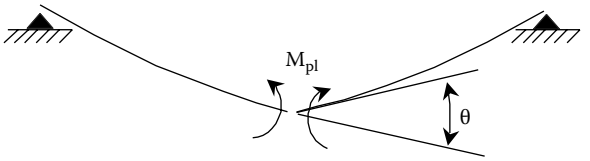


FIGURE 42.22 Plastic hinge bending moment versus rotation angle.

In order to describe the use of the plastic hinge approach, take the structural subsystems represented in Figure 42.21. Each structural component can be modeled as a collection of rigid bodies connected by joints and nonlinear spring elements. Depending on the particular location of the component and on the deformation expected, a choice for each joint type is required and necessary to ensure that the correct deformation mechanism can develop.

In the particular case of a flexural plastic hinge, the spring stiffness is expressed as a function of the change of the relative angle between two adjacent bodies connected by that plastic hinge, as shown in Figure 42.22. For a bending plastic hinge, the revolute joint axis must be perpendicular to the neutral axis of the beam and to the plastic hinge bending plane simultaneously. The relative angle between the two adjacent bodies measured in the bending plane is

$$\theta_{ij} = \theta_i - \theta_j - \theta_{ij}^0 \tag{42.20}$$

where $\theta_{ij}^{\,0}$ is the initial relative angle between the adjacent bodies. Note that for the case of flexible adjacent bodies, the relative angular values also include information on the nodal rotational displacements.

The constitutive relation of the plastic hinge is represented by a load–deformation curve, such as the one presented in Figure 42.23. These constitutive relations are obtained through analytical, experimental, or computational results. In this work, the methodology of identifying the constitutive laws of the plastic hinges is based on the use of FE models of the structural subsystems. A nonlinear static FE analysis is performed for each structural member to identify the force–displacement curve of each plastic hinge.

The constitutive models identified by static FE analysis do not take into account some of the dynamic effects that are exhibited by structural elements when subjected to relatively fast deformation processes. In fact, some of the materials used in vehicle structures are strain rate sensitive, which requires a modification of the constitutive models identified by static elastic–plastic analysis. Steels exemplify the type of materials that are strain rate sensitive, while aluminums are not. A dynamic correction factor is used to account for the strain rate sensitivity [80]:

$$\frac{P_d}{P_s} = 1 + 0.07\, V_0^{0.82} \tag{42.21}$$

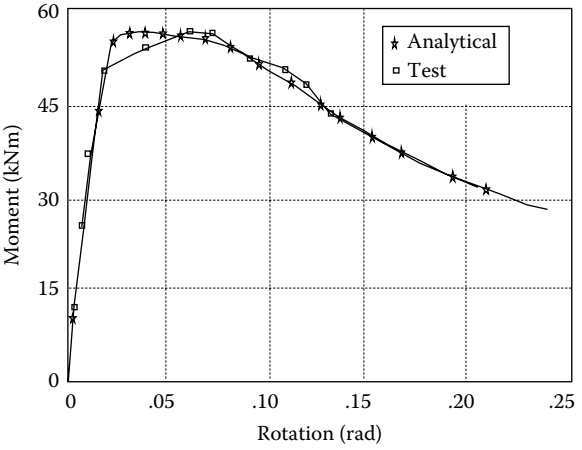


FIGURE 42.23 Plastic hinge bending moment constitutive relationship.

where

P_d and P_s are the dynamic and static forces, respectively

V_0 is the relative velocity between the adjacent bodies

The coefficients appearing in Equation 42.21 are dependent on the type of cross section and material. A detailed description of the application of plastic hinge to crashworthiness can be found on Refs. [66–68].

42.6 DEVELOPMENT OF A VEHICLE MODEL FOR FRONT IMPACT

The construction of a vehicle MB model for front impact is used here not only to exemplify how tests in regulation can be used for vehicle development but also to show that even with a generic description of the vehicle structure and mechanical components, it is possible to achieve models that have a crash response similar to that of an existing vehicle that is not fully known. In this sense the vehicle model developed is designated as a generic vehicle. For the sake of using available resources, the development of the MB generic model presented in this work is based on a previously developed detailed FE model of a small car. The crash response of the FE model is used as reference for the final development of the MB model. Note that the existence of the detailed FE model is not mandatory for the development of the MB model. A vague geometric description of the vehicle structural subcomponents and suspension systems or even the real vehicle would suffice for this development. In this case, the reference crash response required for the final development of the MB model could be the experimental data of the real vehicle, if it exists, or published data from ECE, NHTSA, or EuroNCAP tests of vehicles of the same class or even fictitious crash response data with the behavior targeted for the vehicle. The model presented is part of the European project APROSYS-SP7 [81].

The RADIOSS [51] FE model, shown in Figure 42.24, is a generic car model of a nonexisting vehicle developed by Centro Ricerche FIAT (CRF) [82]. In the MB model of the vehicle, only the structural impact-resistant components are considered. Such structural components are represented in the vehicle shown in Figure 42.24. In a side impact crash application, the door panels are extremely important, and, therefore, they are also included in the model. In a frontal impact, all structural elements are important from the front of the vehicle up to the B-pillar. In both side and front impacts, the rear of the vehicle does not play a role, and, therefore, its structural components are not considered as such but are still included as rigid elements to account for their mass and inertia.



FIGURE 42.24 FE model of a generic car model. (From Puppini, R. et al., Generic car (FE) models for categories super minis, small family cars, large family executive cars, mpv and heavy vehicle, Technical Report APROSYS AP-SP7-0029-A, 2005.)

The first part of the procedure consists in using the geometry of the FE vehicle model and of its structural components to develop the geometry of the MB vehicle model, so that both models have the same dimensions. It is important to underline the importance of correct dimensioning of the vehicle since it has a crucial importance in all deformation processes including intrusions, mechanisms of deformation, and energy absorption. The second part of the procedure consists in dividing the vehicle components in substructures that can be independently modeled by "strings" of rigid bodies connected by plastic hinges. In each of the substructure models, there are specific plastic hinge constitutive equations added to each one of the kinematic joints that connect the rigid bodies in the "string" [66–68]. These constitutive equations are obtained through the use of FE substructural models, deformed with suitable loading cases using ABAQUS FE [83]. The final modeling step of the procedure consists in the MB submodel assembly. The MB code MADYMO [48] is used to represent the model developed as it allows the representation of the plastic hinges and all forms of constitutive equations by point and Cardan restraints. The complete MB vehicle model for frontal impact crash is presented in Figure 42.25.

The MB generic car model is divided into fifteen subsystems, presented in Table 42.9; the vehicle chassis model substructures and all equivalent MB subsystem models are represented. Among these

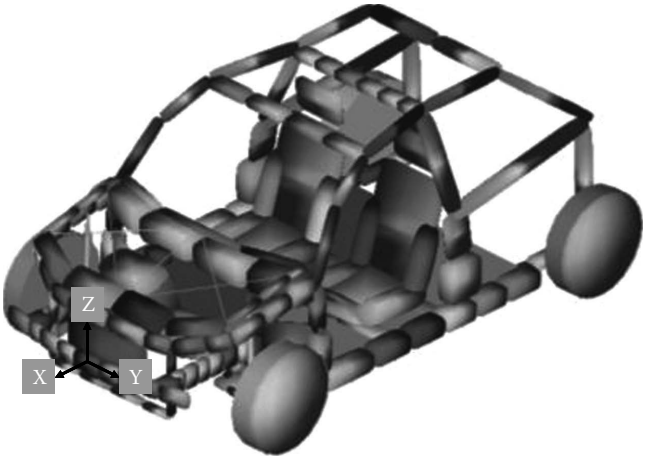
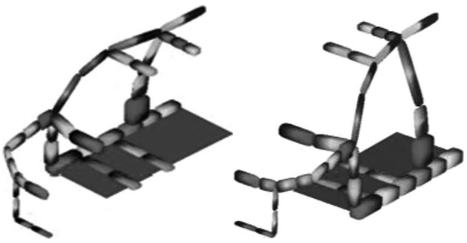
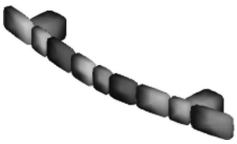


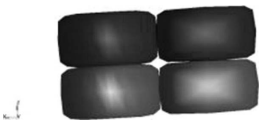
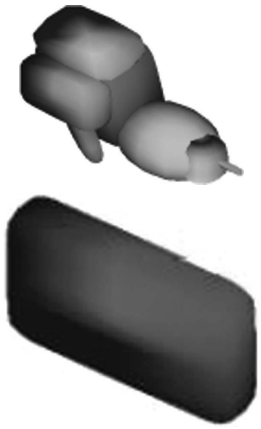
FIGURE 42.25 MB model of the generic car.

TABLE 42.9
Subsystems for Vehicle Structural and Mechanical Components

FE Submodels	MB System Models	Description
		Body right-hand side Body left-hand side
		Rear
		Frame
		Bumper
		Left/right suspension Rear left/rear right wheel

(*continued*)

TABLE 42.9 (continued)
Subsystems for Vehicle Structural and Mechanical Components

FE Submodels	MB System Models	Description
		Hood
		Left door/right door
		Engine/radiator
		

elements, the front and rear suspension systems and wheels are also included. Also the interaction between the road surface and the tire is included in the models through the use of a suitable tire model. Because the emphasis of the development of the model is on the structural components and not on other mechanical components of the vehicle, no more details on the suspension systems are presented here.

The rear of the vehicle does not play a role in any of the crash cases for which the MB model is developed. Figure 42.26 represents the rear of the vehicle as a collection of rigid bodies connected by bracket joints, that is, kinematic joints that restrain any relative degree of freedom between the bodies they connect. Thus, for computational purposes, the whole rear vehicle subsystem is considered as a single rigid body that has the mass and inertia properties of the same substructure of the reference vehicle. It should be noted that for the model of the rear of the vehicle, a single rigid body would suffice. If it is necessary to apply the MB vehicle model for rear crashes, the bracket joints

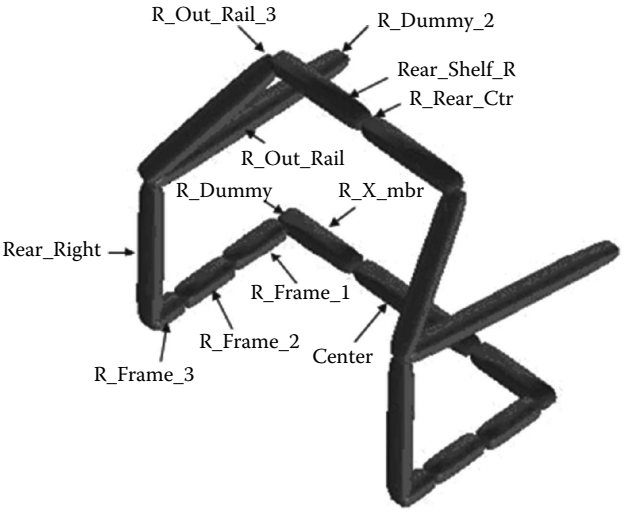


FIGURE 42.26 Vehicle rear subsystem assembly.

can be substituted by other kinematic joints and constitutive equations can be associated to them completing in this form the plastic hinge modeling that enables the subsystem MB model to be used for rear crash analysis.

The constitutive equations for the plastic hinges used to represent the deformation of the vehicle structures are identified here through nonlinear FE static analysis of each of the structural components for different loading cases. Although any suitable nonlinear FE code can be used for this purpose, the FE code ABAQUS is used here for the nonlinear static analysis. It must be noticed that such constitutive behavior may be defined as a design objective or obtained by experimental testing of structural subcomponents. The structural characteristics of the plastic hinges are identified using the explicit method in ABAQUS. The first step to obtain an FE model to calculate the plastic hinge equation is to isolate the part of the FE model of the substructure where the plastic hinge is located. Figure 42.27 shows the FE model of the generic car bumper in ABAQUS [83]. For the material plasticity law, a Johnson–Cook constitutive model is selected [84].

Figure 42.28 represents a piece model of the bumper frame that is used to evaluate the plastic hinge constitutive equation. The plastic hinge presented here is located in the right side of the model. This plastic hinge is associated to one or more kinematic joints in the bumper frame system model [81]. The nonlinear FE analysis required to obtain the plastic hinge constitutive equations consists in fixing one side of the model and applying moments or traction/compression forces on the other side. In the MB model of the bumper, moments are applied to identify the constitutive equation of the plastic hinge and a revolute joint is used to describe the mechanism of deformation of this hinge. Only a bending moment about each axis is applied to find the constitutive behavior of

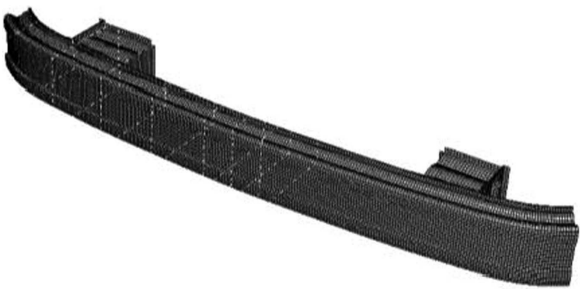


FIGURE 42.27 Bumper FE model in ABAQUS.

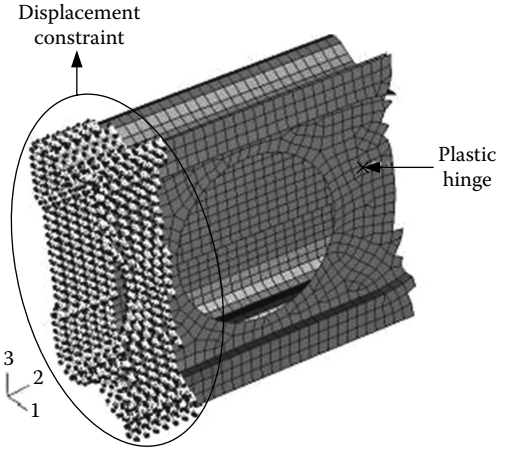


FIGURE 42.28 Boundary conditions for the loading of the bumper frame system model.

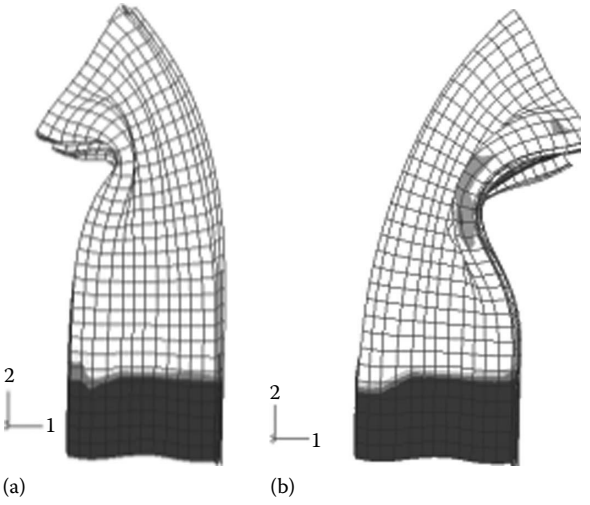


FIGURE 42.29 Bumper deformations for moments applied in counterclockwise (a) and clockwise (b) directions.

the hinge in each direction. The plastic hinge constitutive equation reflects the response of the substructure under the loading case described. Figure 42.29 presents the behavior of the plastic hinge as a deformation of the structure clockwise and counterclockwise moments

The function obtained for the loading case is presented in Figure 42.30. It results from applying bending moment that leads to rotations in the interval [−1, 1] radians to the FE model of the substructure element. The graph shown in Figure 42.30 cannot be directly implemented in the MB model. The constitutive behavior of the plastic hinge is approximated by a piecewise linear function during a loading and an unloading path, representing elastic unloading, parallel to the first segment of the function. The number of segments that are used to represent the load–deformation curve should be small and without large oscillations in the loading path.

Note that the process must be repeated for all plastic hinges considered in the model of the vehicle. The elastic deformations that are observed in the FE model of the structural components between the plastic hinges are not represented in the MB model. This leads to an increased stiffness of the MB structural models that can be compensated by artificially decreasing the stiffness of the

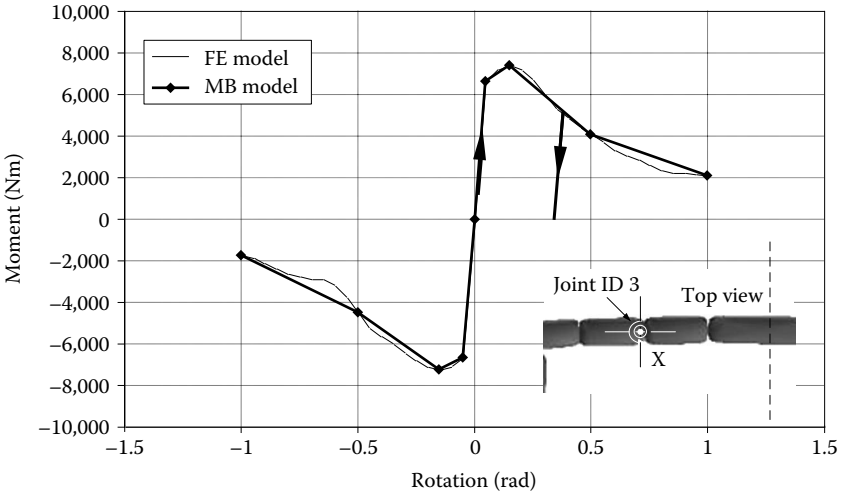


FIGURE 42.30 Constitutive relation for the plastic hinge defined by the revolution about joint *Z* axis.

plastic hinges. Such corrective measures are applied during the validation (correlation) of the MB model with the reference crash behavior and not at this modeling stage.

The MB model of the vehicle is simulated in a crash test that corresponds to the FMVSS208 front crash with a rigid barrier. The vehicle model is used as developed using the procedure outlined, and, therefore, it is not yet validated, that is, there is no assurance that its crash response matches a reference crash behavior. The crash response of the vehicle can be measured by accelerations and velocities in selected points of the structure or by evaluating the energy absorption during the crash event. For vehicle models equipped with dummies, which is not the case in the current analysis, measures of injury of different body parts are relevant not only to describe the quality of the model but also for validation purposes.

The validation process, with respect to the reference behavior obtained from the equivalent FE model [82], consists in modifying the plastic hinge constitutive relations, criteriously, to have a better correlation between the reference FE vehicle model and the MB model responses. Through a trial and error procedure, the plastic hinges of the different parts of the vehicle have been adjusted. The model obtained is tested for the same case as the original vehicle and the outcome is displayed in Figures 42.31 through 42.33 that show the crash response of the validated vehicle. The acceleration signals are filtered with a CFC 60 filter.

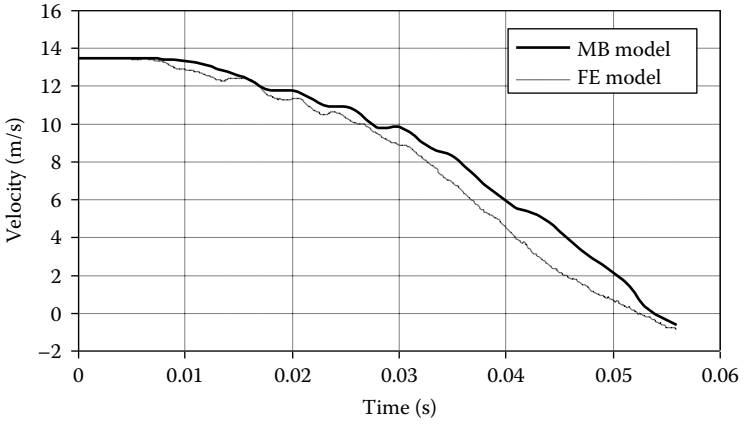


FIGURE 42.31 Left B-Pillar above sill velocities for the vehicle in frontal impact.

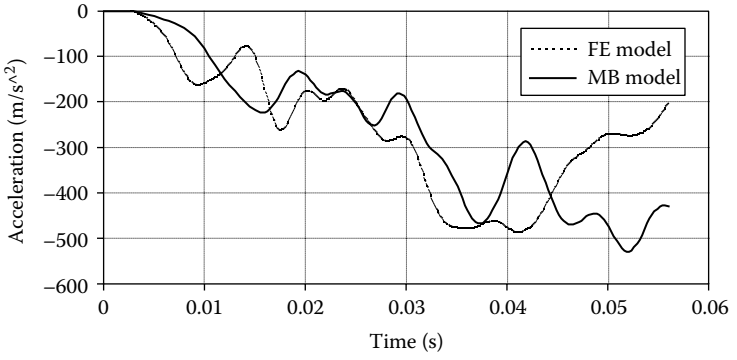


FIGURE 42.32 Left B-Pillar above sill accelerations of the vehicle in frontal impact.

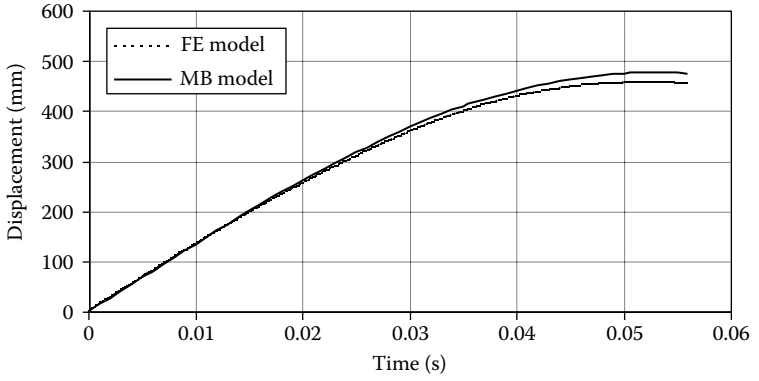


FIGURE 42.33 Left B-Pillar displacements of the vehicle in frontal impact.

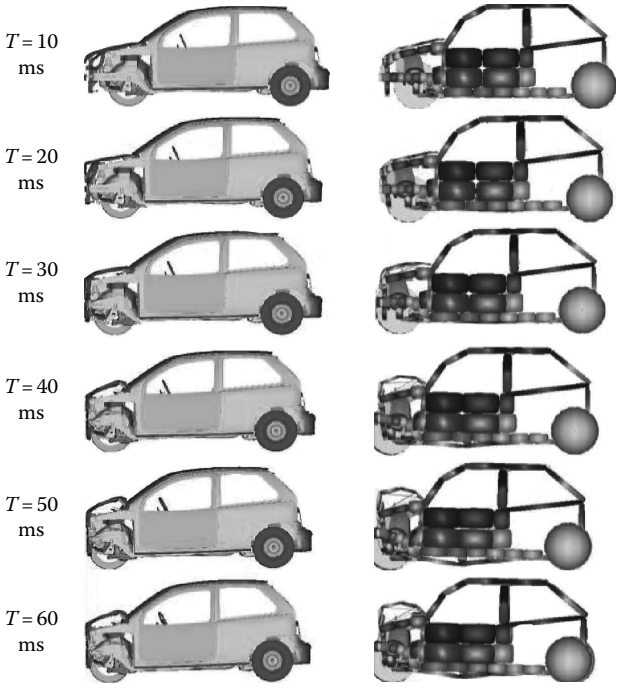


FIGURE 42.34 MB and FE models deformations in the frontal crash test.

The validated model correlates well with the response of the reference model. The pattern of deformation that both models exhibit, shown in Figure 42.34, also demonstrates that the mechanisms of energy absorption are similar in both models. Although the correlation represented by Figure 42.34 is purely qualitative, quantitative results for the correlation of the energy absorption of the different subsystems have also been obtained, and they confirm the results presented.

42.7 DEVELOPMENT OF SIDE IMPACT PROTECTION SYSTEMS

The aim of the study described here, part of project APROSYS-SP6 [85], is to devise an actuator that can be deployed about 200 ms before the actual crash starts that improves the occupant protection. Due confidentiality reasons, the procedure presented is used to evaluate the seat repositioning and door interior trimming instead of the system that has finally been achieved. An MB model of a vehicle Chrysler Neon, shown in Figure 42.35, developed according to the methodologies described previously is analyzed in the crash scenario described in the regulation FMVSS 214 for the purpose of devising a precrash deployment system that helps improve the occupant protection in a side impact. The analyses of the MB models are carried using the code MADYMO [48]. The starting point is the MB model for the Chrysler Neon in which the actuator has to operate.

The kinematic joints represented in Figure 42.36 are set according to the expected mechanisms of deformation that the vehicle can experience for the side crash tests. Caution must be exercised if the model is to be used in any other type of side impact test besides the ones prescribed in the FMVSS 214 or ECE 95 regulations. It is assured that not only the regions of the vehicle for which deformations can happen are properly described in the model but also the potential pairs of contact are defined.

The energy absorption of the vehicle side structure in the test configuration can only be realized when it is included in the complete vehicle. Because all structural plastic deformations are contained in the domain of the side structure, the remaining of the chassis can be considered rigid. The attachments between the rigid bodies of the side structure and the rigid body of the chassis are realized by the crushable elements. The plastic deformation energy that these elements account for is due to the deformation of the structure surrounding the roof, sill, A-pillar, and C-pillar. If the deformation of these crushable elements exceeds a given limit, it can be argued that the plastic deformations of the side structure exceed the modeled region of the vehicle, and, therefore, the validity of the model can be questioned.

Figure 42.37 shows the results obtained for the simulation of the FMVSS 214 side crash test in terms of head, pelvis, thoracic vertebras, and midrib accelerations. The U.S.-SID injury criteria corresponding to this simulation are presented in Table 42.5. The value of TTI is above the maximum acceptable value specified in FMVSS 214 and the value for the PLA is also quite high but below the limit specified in the norm. The velocities of different points of the structure and dummies match those obtained in the experimental test.

The contact forces for the dummy presented in Figure 42.38 have not been measured in the experimental test. It is observed that the peak forces for the pelvis, thorax, and head occur simultaneously. The pelvic contact force predicted by the simulation is clearly larger than the forces carried by the head and by the thorax. This is mainly due to the proximity of the seat and occupant driver

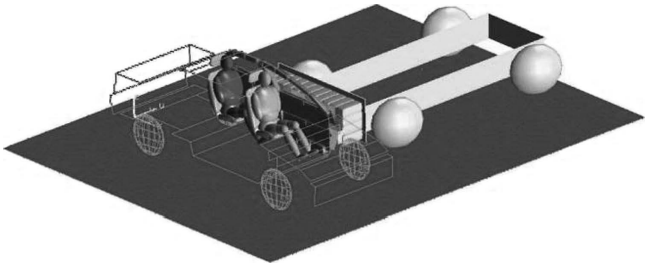


FIGURE 42.35 MB model for the Chrysler Neon in the side impact test defined by FMVSS 214.

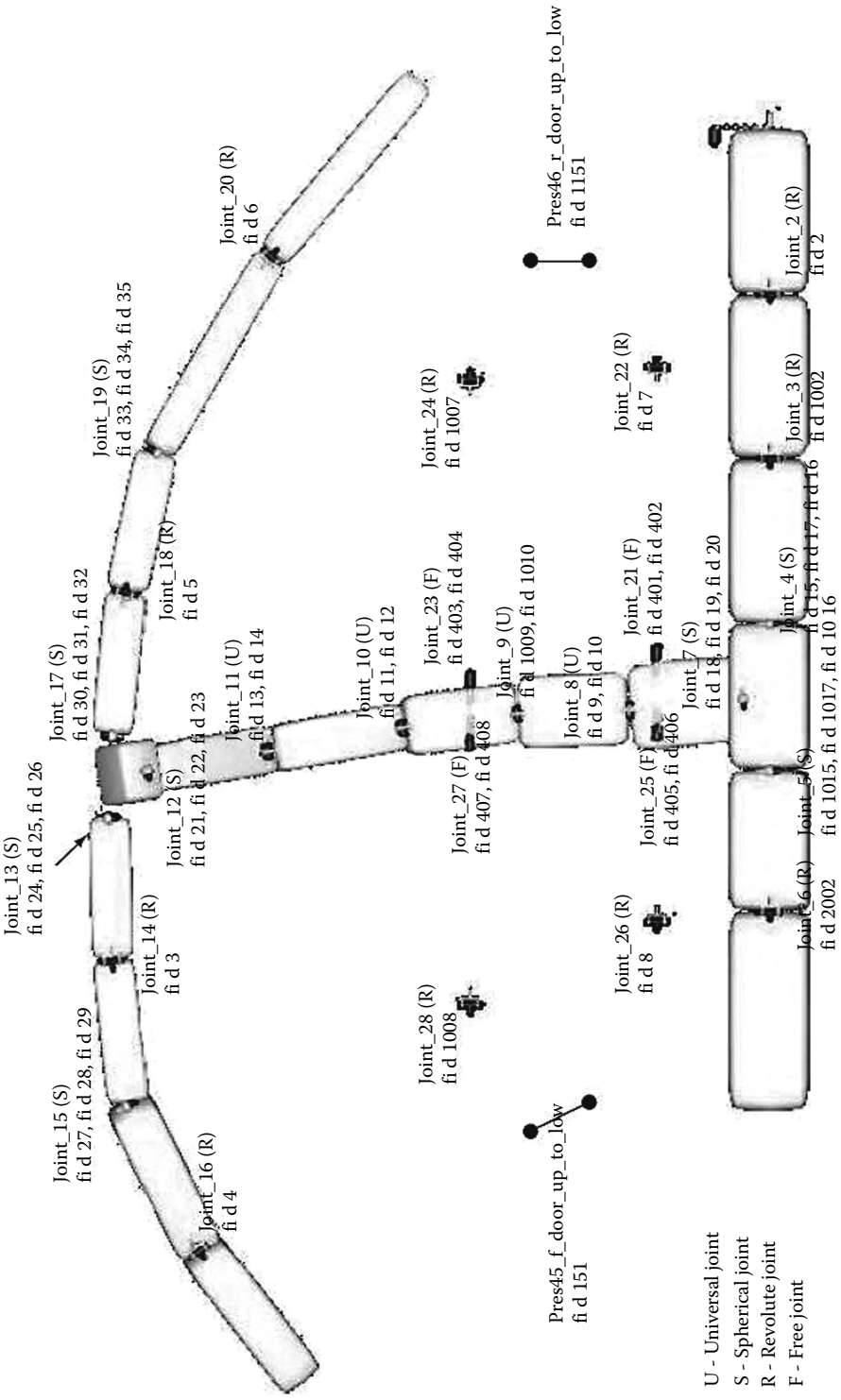


FIGURE 42.36 Kinematic joints for the side structure representation.

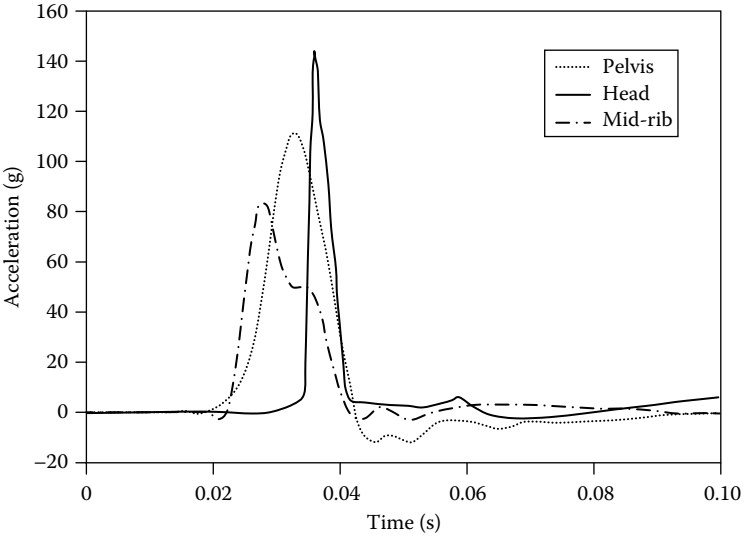


FIGURE 42.37 Side component accelerations for parts of the dummy during impact.

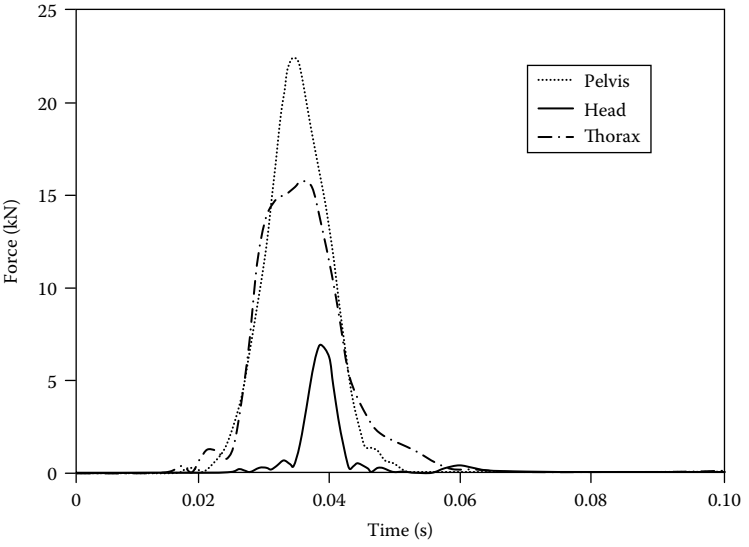


FIGURE 42.38 Contact forces in the pelvic region, thorax region, and head.

pelvis to the B-pillar, which is directly struck by the barrier. The B-pillar motion and the occupant kinematics are displayed in Figure 42.39 for different instants of time. It is clear that the B-pillar intrudes the vehicle passenger compartment, with special incidence in its lower part. The intrusion of the B-pillar leads to contact with the driver's seat and pushes it to the vehicle interior.

The results of the simulation also show the importance of modeling the suspension of the vehicle and the contact between the tires and the ground. The vehicle roll and its lateral displacement play an important role in the progression of the crash event. The energy dissipated by the suspension and friction between tires and ground cannot be neglected for lower impact speeds.

The results for the side impact demonstrate the reliability of the MB model of the vehicle and occupants used in this crash scenario. Extensive reanalysis of the complete vehicle is only possible when the computation time for each analysis is low. To appraise the computational efficiency of this methodology, the time required for each simulation of the vehicle side impact with the MB

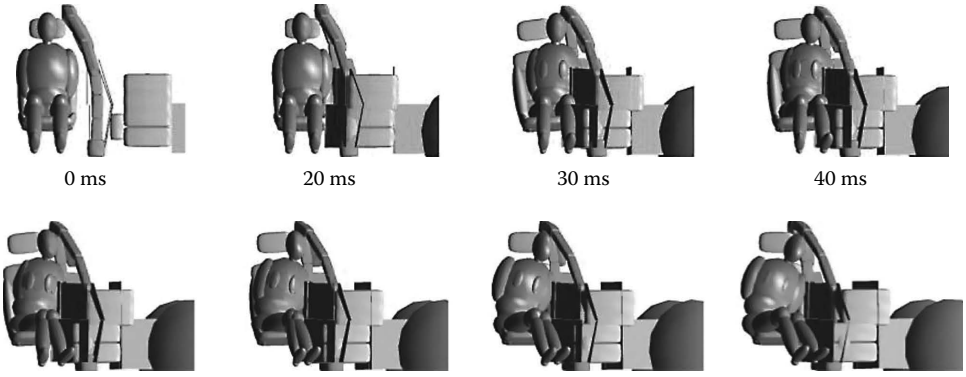


FIGURE 42.39 Sequence of images for the side crash.

formulation is measured in minutes, while the simulation of the equivalent model in FE, for the same test conditions, is measured in days. The evaluation of the benefits of having the driver sitting in a different position is studied in different simulations. The main objective is to find if repositioning the driver's seat in a favorable side crash position should be considered as an effective side protection strategy. Note that the precrash actuation solution used in the APROSYS-SP6 project [85] is different from that reported here, although the methods used are the same.

The first strategy moves the driver sitting positions to the interior of the vehicle by 5 cm, as shown in Figure 42.40, so that the space between the driver and the intruding door is increased. In Figure 42.40c, the dark outline shows the seat and dummy initial positions and the solid dummy drawing shows the seat and dummy positions for the simulation. The repositioning of the occupant and seat further to the interior of the vehicle, by 5 cm, leads to an improvement of the PLA and of the TTI, as shown in Table 42.10. It should be noted that the position of the seat must be followed by the seated occupant, who is assumed to be correctly seated.

Analyses of the seating scenario with small perturbations of the occupant position indicate that the injury criteria are sensitive to position changes of the dummy as well as changes in the pelvic and thorax velocities, caused by the seat movement. This indicates that if the change on the seat position inside the car is to be made in a short period of time, the effective repositioning of the occupant must also to be pursued. Finally, note that the improvements on the injury criteria happen independently of the increase in intrusion registered in the side structure of the vehicle. The increase of intrusion is explained by the removal of the seat to a more interior position in the vehicle reducing in this form the intrusion resistance.

In the second scenario, the driver sitting position is relocated to the back of the vehicle by 15 cm, as shown in Figure 42.41, to be seated in a region of the vehicle exposed to reduced intrusion

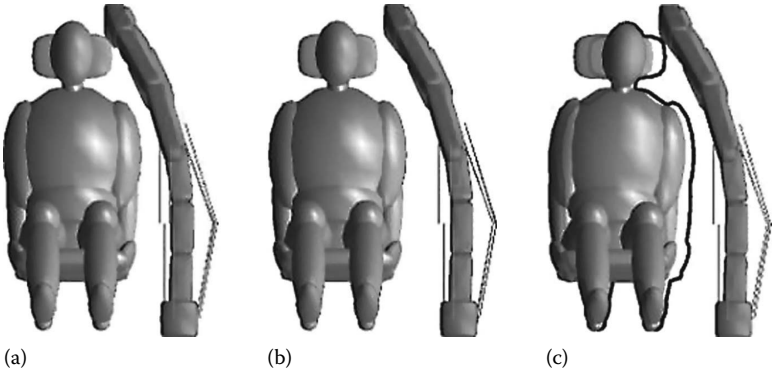
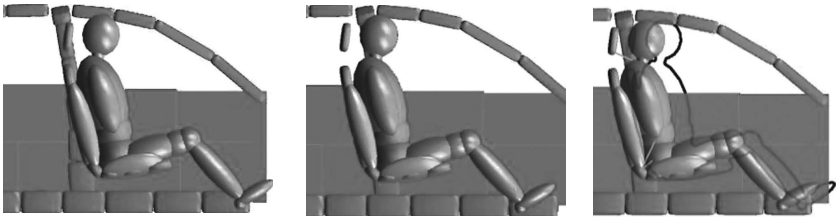


FIGURE 42.40 Driver sitting position: (a) original; (b) located 5 cm to the interior of the target vehicle; (c) difference in the seating positions.

TABLE 42.10
Injury Criteria and Their Variations for the Driver Lateral Translation

	Injury Criteria Value	
Sideway	**PLA (<130 g)**	**TTI (<85 g)**
Initial design	110.23	87.83
Driver seat 5 cm to interior	81.13	77.16
Change in injury index (%)	−26.4	−12.15



(a) (b) (c)

FIGURE 42.41 Driver sitting position: (a) original; (b) further back by 15 cm in the target vehicle; (c) difference in the seating positions.

displacement and speed. Placing the seat more to the back of the vehicle is more effective at pelvic level, which is probably due to the fact that intrusion and intrusion velocity values are lower closer to the B-pillar than in the region corresponding to the middle of the door.

The results presented in Table 42.11 indicate that some improvement of the dummy injury criteria can be obtained by moving the seat backward in the vehicle prior to the crash. The decrease in the PLA is consistent with what was previously determined in simulation. The differences between the magnitudes of the decrease of the PLA are due to the final positions of the dummy relative to the horizontal intrusion profile in the car side structure. A large difference is observed for the TTI in the simulation. This is due to the fact that for simulation, the distance between the thorax of the dummy and the door is altered due to the seat belt restraint.

Pulling the driver's seat backward prior to the occurrence of a side crash improves the protection level. However, a more detailed analysis of the model must be undertaken to ensure that not only the seat is moved but also that the occupant moves with it. The possible benefits of rotating the driver's seat in the target vehicle in such a way that the driver is displaced more to the interior of the vehicle, as seen in Figure 42.42, are investigated here also.

In order to ensure that the occupant moves with the seat to the precrash position, a seat-mounted seat belt is used to improve the coupling of the dummy's torso movement with that of the rotating seat.

TABLE 42.11
Injury Criteria and Their Variations for the Driver Forward Translation

	Injury Criteria Values	
Forward–Rearward	**PLA (<130 g)**	**TTI (<85 g)**
Initial design	110.23	87.83
Driver seat moved back 15 cm	73.76	41.87
Change in injury index (%)	−33.08	−52.33

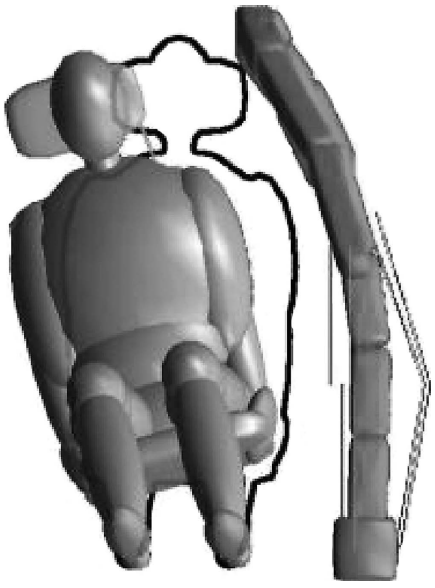


FIGURE 42.42 Rotation of the driver's seat prior to the side impact.

If a protection strategy involving the movement of the driver's seat is to be implemented, further analysis of this model, using a seat belt model that can slide on the dummy, has to be used. Furthermore, some additional devices might be required to assure that the torso movement follows that of the seat.

Table 42.12 shows the results for TTI and the PLA measured in the side direction. The results presented in Table 42.12 show that the rotation of the seat in the direction of the interior of the target vehicle can be a very efficient protection strategy in the event of a side crash. In fact, reductions of up to 45% and 50% are obtained in the PLA and TTI if the seat is rotated by 16°.

The limitation of the intrusion of the door and B-pillar allows for the space between the occupant and the interior of the door to be used for the lateral excursion of the occupant in case of a side impact. The 3-point seat belts and the seat shape are not effective restraints for this occupant side motion. Therefore, the interior of the vehicle, and the door in particular, must have a compliance that helps in reducing injury when it is hit by the occupant during a side impact [86,87]. The part of the study described here exemplifies the use of the numerical tools to develop the functional requirements for a deployable/retractable equipment that can not only limit the lateral excursion of the occupant but also increase the interior compliance of the vehicle for side impact scenarios.

TABLE 42.12
Injury Criteria Forces and Their Variations for the Driver's Seat Rotation

	Injury Criteria Values	
Seat Rotation	**PLA (<130 g)**	**TTI (<85 g)**
Initial design	110.23	87.83
Injury criteria	61.64	42.95
Change in criteria (%)	−44.08	−51.1
	Pelvis	**Thorax**
Contact force change (%)	−30.05	−57.53
Resultant acceleration change (%)	−43.11	

A set of simulations is carried out to evaluate the side crash protection potential of having an improved cushioning for the pelvis and thorax consisting of specifically designed crush pads, that is, pads in the interior of the door that can occupy the space in between the occupant and the door as schematically shown in Figure 42.43a. For this purpose, two crush pads are modeled in order to provide independent cushioning of the thorax and pelvis, as seen in Figure 42.43a. These cushioning devices were inserted into the intruding door in a manner that all the free space between the door and the dummy can be used for cushioning and thus to contribute to the side crash performance improvement. It is considered that each of the pads behaves according to the force–deflection curves presented in Figure 42.43b. Note that the values for the pad stiffness are fictitious being only used here to demonstrate the concept.

The simulation where no damping is included in the results is reported in Table 42.13 for reference. Note that the model with no damping is rather unrealistic, as there all materials have some internal damping. The results presented in Table 42.13 show that for "soft" and for "stiff" pad materials, the amount of internal damping present in the material plays an important role in the side crash outcome.

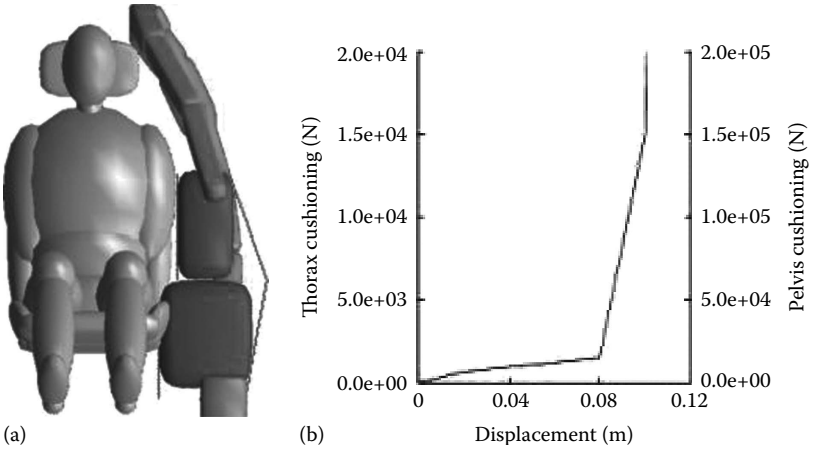


FIGURE 42.43 (a) Location for the door cushioning, that is, crash pads at the pelvic and thorax level; (b) force–displacement curve describing the material behavior of the cushioning pad.

TABLE 42.13
Injury Criteria Values for Different Values of the Pad Material Damping

	Injury Criteria Values	
Door Interior Padding	**PLA (<130 g)**	**TTI (<85 g)**
Initial design	110.23	87.83
No damping	82.77	83.29
Low damping	79.55	75.68
Medium damping	69.21	66.06
High damping	72.82	80.36
	Percentage Change in Injury Criteria Values	
Door Interior Padding	**PLA**	**TTI**
No damping	−24.92	−5.17
Low damping	−27.83	−13.84
Medium damping	−37.21	−24.79
High damping	−33.94	−8.51

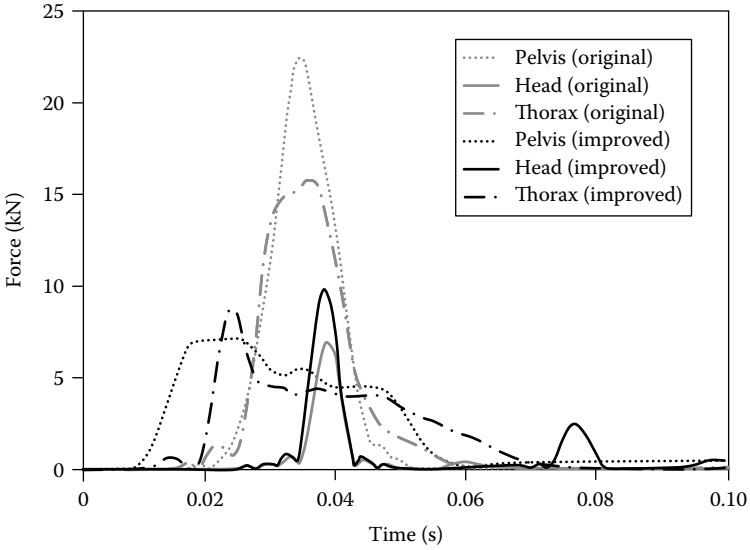


FIGURE 42.44 Dummy contact forces for the initial design and improved designs.

The existence of some internal damping in the pad material is beneficial in what the injury criteria of the driver are concerned. The material damping contributes to the decrease of the contact forces within reasonable values, but if this damping assumes a high value, the contact forces increase again.

Figures 42.44 and 42.45 illustrate the improvements obtained by the use of the optimized crush pads. The peak forces in the order of 25 kN, at the pelvic level, observed for the original design are now reduced to about 10 kN. The lack of compatibility intrusion profile with the occupant rotation that leads to head contact with high forces in the original vehicle is compatible for the new design due to a more upright position of the occupant and more compliance in the contact areas.

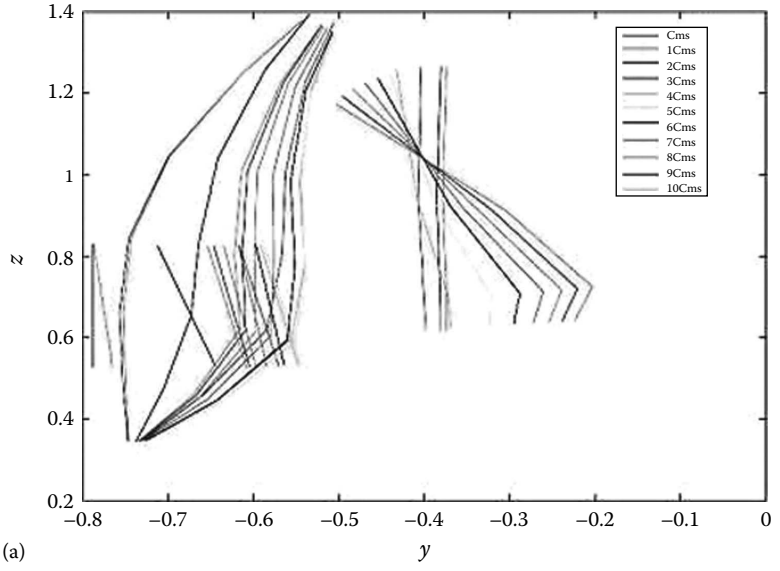


FIGURE 42.45 Intrusion and dummy displacement: (a) initial design.

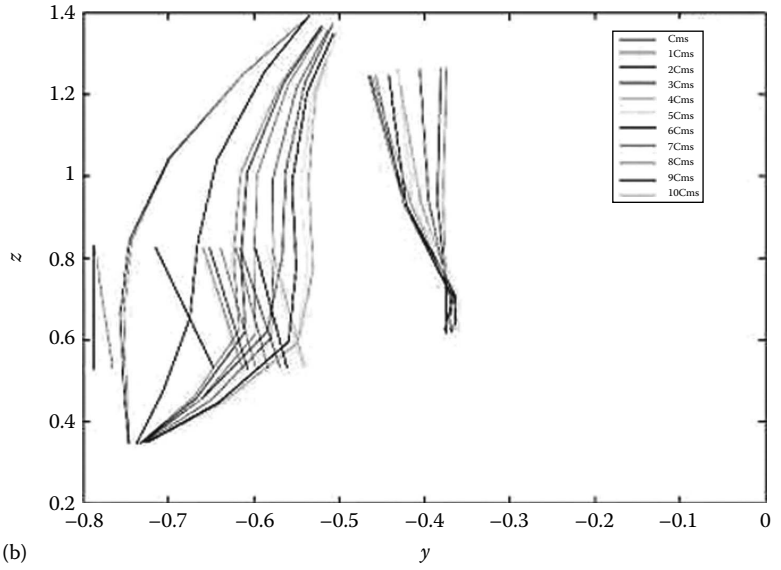


FIGURE 42.45 (continued) Intrusion and dummy displacement: (b) improved design.

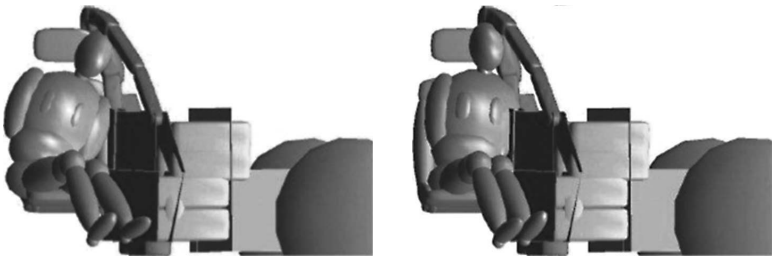


FIGURE 42.46 Dummy displacement at 80 ms for (a) initial design; (b) improved design.

The occupant position at the instant of contact with the side structure of the vehicle for the initial vehicle and for the new design is presented in Figure 42.46. Even the qualitative observation of both cases allows concluding that a more serious pelvic and thorax impact exists for the original design of the vehicle.

42.8 CONCLUSIONS AND TRENDS IN FUTURE DEVELOPMENTS

Occupant protection in automotive vehicles has improved greatly during the past decades as a result of concurrent efforts that integrate vehicle structural solutions with interior occupant protection systems. The strategic lines for this continuous evolution in crashworthiness are defined not only by accidentology analysis, which allows to identify the priorities for occupant protection, but also by technological developments in sensors, sensing algorithms, and interior equipment, which widen the range of decisions to be taken during a crash event, and even before it starts, and enable fitting the characteristics of the protection systems to the individual car occupant. The regulations used for crashworthiness are mostly regional, taking into account the different automobile markets and accident profiles.

Numerical methodologies play a fundamental role on the advancement of vehicle passive safety. Not only the computational mechanics aspects are in constant development, providing more reliable material models, better tools to model structures and systems, faster and reliable algorithms for the

solution of the dynamic equilibrium equations of the vehicles and their subsystems, and improved contact algorithms, but also they are also supported by advanced graphical interfaces, which allow for better look and feel and for defining more detailed and precise information. The constant improvements in computer power ensure that the models developed can be more advanced, better integrated, and more predictable in a reasonable time. The actual difference between FE-based methods and MB dynamic methods has the tendency to diminish as these procedures will be more integrated. Finally, the full integration of other numerical methodologies, such as optimization tools and real-time control paradigms, will increase the importance of the computational methods on the vehicle design.

The development of more advanced knowledge in the human mechanisms of injury; in the modeling of human tissues, organs, and structures; and in the physiologic response to impact will certainly impact the current regulations, the design of the currently used occupant protection systems, and the design of the vehicle structures. Today only ATDs are used in regulations and in internal norms of different companies. Other human occupant surrogates are used, mostly for impact biomechanics research. The test dummies are designed to collect data for prescribed directions of loading and do not include any representation of muscle activation and, consequently, cannot represent realistic human behavior for general crashes. The research on human body modeling is expected to have an impact on the development of more advanced dummy designs and eventually lead to the realization of omnidirectional dummies that include some level of muscle bracing when used in lower impact velocities. Also, advances in the definition of injury criteria are expected as a result of advances in human body modeling. Certainly the regulations will continue to be regional, at least while the car markets are different, but the measuring devices are expected to be unified.

Other aspects of automotive passive safety in which major developments are expected concern all those that can benefit from the integrated use of the many sensor devices existing in the vehicles today. Not only these devices allow for adjusting the deployment of the interior protection devices to the particular characteristics of the crash and of the occupant, but also they already allow for precrash actions that prepare the vehicle and the occupant for a foreseeable crash. The development of new materials and structural solutions will extend the range of the precrash systems from the level of the systems to the actual vehicle structure. Finally, it is expected that many of the aspects of active and passive safety are more and more integrated so that all phases of the vehicle operation can take advantage of common resources, leading to a more integrated vehicle safety.

ACKNOWLEDGMENTS

The support of EC through project TIP3-CT-2004-506503 with the subprojects APROSYS-SP6, with the partners SIEMENS VDO (Ge), Faurecia (Ge), DaimlerChrysler (Ge), CIDAUT (Sp), Fh G (Ge), Warsaw University of Technology (Pl), and TNO (Nl) that provided the input and discussion on vehicle side impact protection, and APROSYS-SP7, with the partners Mecalog (Fr), CRF (It), TNO (Nl), Polytechnic of Turin (It), CIDAUT (Sp), and Graz University of Technology (At) that provided the reference model for the frontal impact study case.

REFERENCES

1. DeHaven, H.: *Mechanical Analysis of Survival in Falls from Heights of Fifty to One Hundred and Fifty Feet*. War Medicine Accident Research Methods and Approaches, New York, 1942, Vol. 2, pp. 586–596.
2. Stapp, J., and Gell, C.F.: Human Exposure to Linear Declarative Force in the Backward and Forward Facing Seating Position. Military Surgeon 109(2), (1951), 106–109.
3. *Proceedings of the 8th Stapp Car Crash Conference*, Way State University Press, Detroit, MI, 1966.
4. Mackay, M.: A review of the biomechanics of impacts in road accidents. In: J. Ambrósio, M. Pereira and F. Pina da Silva (eds.): *Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997, pp. 115–138.

5. Schmitt, K.-U., Niederer, P., and Walz, F.: *Trauma Biomechanics: An Introduction to Accidental Injury*. Springer, Heidelberg, Germany, 2004.
6. Brevet d'Invention no 331.926. Republique Francaise. Office National de la Propriete Industrielle, 1903.
7. Adams, J.: *The Efficacy of Seat Belt Legislation*. SAE Technical Paper 820819, SAE Transactions, SAE International, Warrendale, PA, 1982, pp. 2824–2834.
8. Janssen, W.: Seat-belt wearing and driving behavior: An instrumented-vehicle study. *Accid Anal Prev* 26(2), (1994), 249–261.
9. Barényi, B.: Kraftfahrzeug, Insbesondere zur Beförderung von Personen, Deutsches Patentamt Nr. 854157 30, 1952.
10. Seiffert, U. and Wech, L.: *Automotive Safety Handbook*. SAE International, Warrendale, PA, 2003.
11. Kokkula, S., Langseth, M., Hopperstad, O., and Lademo, O.-G.: Offset impact behavior of bumper beam-longitudinal systems: Experimental investigations. *Int. J. Crashworthiness* 11(4), (2006), 299–316.
12. Klanner, W.: Status report on future development of EuroNCAP programme. *Proceedings of the 17th ESV Conference*. Amsterdam, the Netherlands, 2001.
13. Wismans, J.: Introduction in injury biomechanics. In: J. Ambrósio, (ed.): *Crashworthiness: Energy Management and Occupant Protection*. Springer, Vienna, Austria, 2001, pp. 305–331.
14. *The Abbreviated Injury Scale, 1990 version/update 98*. Association of Advancement of Automotive Medicine. Dea Plaines, IL, 2001.
15. Viano, D.C. and King, A.I.: Injury mechanisms and biofidelity of dummies. In: J. Ambrósio, M. Pereira, and F. Pina da Silva (eds.): *Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997, pp. 25–51.
16. Versace, J.: A review of the severity index. *Proceedings of the 15th STAPP Car Crash Conference*, SAE, New York, 1971, pp. 771–796.
17. Lissner, H.R., Lebow, M., and Evans, F.G.: Experimental studies on the relation between acceleration and intracranial pressure changes in man. *Surg. Gynecol. Obstet*. 111, (1960), 329–338.
18. Newman, J.A.: A generalized acceleration model for brain injury threshold. *Proceedings of the International Conference on Biomechanics of Impact 1986*. IRCOBI, Lyon, France, 1986, pp. 121–131.
19. Wismans, J., Janssen, E.G., Beusenberg, M., Koppens, W.P., and Lupker, H.A.: *Course Notes in Injury Biomechanics*. Faculty of Mechanical Engineering, Technical University of Eindhoven, Eindhoven, the Netherlands, 1994.
20. Boström, O., Svensson, M., Aldman, B., Hansson, H., Håland, Y., Lövsund, P., Seeman, T., Suneson, A., Säljö, A., and Örtengen, T.: A new neck injury criterion candidate based on injury findings in the cervical spinal ganglia after experimental neck extension trauma. *Proceedings of the International Conference on Biomechanics of Impact 1996*. Lyon, France, 1996, pp. 123–136.
21. Kleinberger, M., Sun, E., Eppinger, R., Kuppa, S., and Saul, R.: *Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems*. NHTSA Report, Washington, DC, 1998.
22. Schmitt, K.-U., Muser, M., Walz, F., and Niederer, P.: N_{km}—A proposal for a neck protection criterion for low-speed rear-ends impacts. *Traffic Inj. Prev*. 3(2), (2002), 117–126.
23. Panjabi, M., Wang, J., and Delson, N.: Neck injury criterion based on intervertebral motions and its evaluation using an instrumented neck dummy. *Proceedings of the International Conference on Biomechanics of Impact 1999*. IRCOBI, Lyon, France, 1999, pp. 179–190.
24. Viano, D. and Davidsson, J.: Neck displacement of volunteers, BioRID P3 and hybrid III in rear impacts: Implications to whiplash assessment by neck displacement criterion (NDC). *Proceedings of the IIWPG/IRCOBI Symposium*. Isle of Man, U.K., 2001.
25. Heitplatz, F., Sferco, R., Fay, P., Kim, A. and Prasad, P.: An evaluation of existing and proposed injury criteria with various dummies to determine their ability to predict levels of soft tissue neck injury seen in real world accidents. *Proceedings of the 18th ESV Conference*. Nagoya, Japan, 2003.
26. Tarriere, C., Walfisch, G., and Fayon, A.: Synthesis of human tolerances obtained from lateral impact simulations. *Proceedings of the 7th ESV Conference*. NHTSA, Washington, DC, 1979, pp. 359–373.
27. Viano, D.: Biomechanical responses and injuries in blunt lateral impact. *Proceedings of the 33rd STAPP Car Crash Conference*, SAE, Warrendale, PA, 1971, pp. 113–142.
28. Viano, D. and Lau, I.V.: A viscous tolerance criterion for soft tissue injury assessment. *J. Biomech*. 21(5), (1988), 387–399.
29. Haffner, M.: Synthesis of pelvic fracture criteria for lateral impact loading. *Proceedings of the 10th ESV Conference*. Washington, DC, 1985.
30. Yamada, H.: *Strength of Biological Materials*. R.E. Krieger Publishers, New York, 1970.
31. IIHS: *Frontal Offset Crashworthiness Evaluation Guidelines for Rating Injury Measures*. Insurance Institute for Highway Safety, Arlington, VA. 2004.

32. Seek, A., Coxon, C., and Wani, K.: Global NCAP harmonisation—Initial Position, Requirements and Prospects. *Int. J. Crashworthiness* 8(1), (2003), 5–15.
33. European Council for Europe: Regulation 94–Uniform provisions concerning the approval of vehicles with regard to the protection of the occupants in the event of a frontal collision. 1995.
34. National Highway Traffic Safety Administration: Federal motor vehicle safety standard no 208, Occupant Crash Protection. *Code of Federal Regulations Part §571.208*. (2004) pp. 497–580.
35. www.euroncap.com.
36. European Council for Europe: Regulation 95–Uniform provisions concerning the approval of vehicles with regard to the protection of the occupants in the event of a lateral collision. 1995.
37. National Highway Traffic Safety Administration: Federal motor vehicle safety standard no 214, Side Impact Protection. *Code of Federal Regulations Part §571.214*. (2004) pp. 666–677.
38. National Highway Traffic Safety Administration: Federal motor vehicle safety standard no 201, occupant protection in interior impact. *Code of Federal Regulations Part §571.201*. (2004) pp. 463–483.
39. European Council for Europe: Regulation 32–Uniform provisions concerning the approval of vehicles with regard to the structure of impacted vehicles in a rear-end collision. 1993.
40. National Highway Traffic Safety Administration: Federal Motor Vehicle Safety Standard no 303, Fuel system integrity of compressed natural gas vehicles. *Code of Federal Regulations Part §571.303*. (2002), pp. 743–746.
41. Matsui, Y., Ishikawa, H., and Sasaki, A. Proposal of injury risk curves for evaluating pedestrian femur/pelvis injury risk using EEVC upper legform impactor based on accident reconstruction. *Int. J. Crashworthiness*. 11(2), (2006), 97–103.
42. European Enhanced Vehicle-Safety Committee. Improved test methods to evaluate pedestrian protection afforded by passenger cars. EEVC working group 17 report. 1998.
43. Ardoino, P.: Car crash and safety testing. In: J. Ambrósio, M. Pereira, and F. Pina da Silva (eds.): *Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997, pp. 189–205.
44. SAE 2006 International Conference and Exposition of Digital Human Modeling for Design and Engineering, Lyon, France, July 4–6, 2006.
45. Schonpflug, M., Merten, K., and Werniecke, P. Numerical simulation of human kinematics and injuries in side crash scenarios. In: *Proceedings of the Digital Human Modeling for Design and Engineering Symposium*. Rochester, MI, June, 2004.
46. Lescheticky, T. and Gholami, T.: USSID and EuroSID: Two tools for the side impact simulation of car structures. In: J. Ambrósio and M. Pereira (eds.): *Proceedings of the NATO-ASI on Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection: Volume II—Contributed Papers*. Tróia, Portugal, July 7–19, 1996, pp. 19–34.
47. Wismans, J.: Design tools: Human body modelling. In: J. Ambrósio, (ed.): *Crashworthiness: Energy Management and Occupant Protection*. Springer, Vienna, Austria, 2001, pp. 333–377.
48. TNO Automotive: *MADYMO Manuals, Version 6.2*. TNO MADYMO BV., Delft, the Netherlands, 2004.
49. Hallquist, J.O.: *LS-DYNA Theoretical Manual: Nonlinear Dynamics Analysis of Structures*. Livermore Software Technology Corporation, Livermore, CA, 1998.
50. Haug, E.: The PAM-CRASH Code as an efficient tool for crashworthiness simulation and design. *Proceedings of the Second European Cars/Trucks Simulation Symposium*. Schliersee, Germany, May 22–24. 1989.
51. Mecalog: *RADIOSS Input Manual, Version 4.2*. Mecalog, Antony, France, 2000.
52. Wittemann, W.J.: Improved vehicle crashworthiness design by control of the energy absorption for different collision situations. PhD dissertation. Technical University of Eindhoven, Eindhoven, the Netherlands, 1999.
53. Wågström, L., Thomson, R., and Pipkorn, B.: Structural adaptivity in frontal collisions: Implications on crash pulse characteristics. *Int. J. Crashworthiness* 10(4), (2005), 371–378.
54. Chan, C.-Y.: *Fundamentals of Crash Sensing in Automotive Air Bag Systems*. SAE International, Warrendale, PA, 2000.
55. Cesari, D.: Advanced restraint systems for occupant protection. In: J. Ambrósio, M. Pereira, and F. Pina da Silva (eds.): *Crashworthiness of Transportation Systems: Structural Impact and Occupant Protection*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1997, pp. 173–187.
56. Harms, P., Renouf, M., Thomas, P., and Bradford, M. Injuries to restrained car occupants: What are the outstanding problems? In: *Proceedings of the 11th ESV Conference*. SAE Paper no 876029. SAE International, Warrendale, PA, 1987, pp. 183–201.
57. Miller, M.A.: The biomechanical response of the lower abdomen to belt restraint loading. *J. Trauma*, 29(11), (1989), 1571–1584.

58. Wiklund, C. and Larsson, H.: *Saab Active Head Restraint (SAHR)—Seat Design to Reduce the Risk of Neck Injuries in Rear Impacts*. SAE Paper no 980297, SAE Transactions, SAE International, Warrendale, PA, 1998.
59. Lundell, B., Jakobsson, L., Alfredsson, B., Lindstrom, M., and Simonsson, L.: The WHIPS Seat—A car seat for improved protection against neck injuries in rear-end impacts. In: *Proceedings of the 16th ESV Conference*, Paper no. 98-S7-O-08, Washington, DC, 1998.
60. Schmitt, K.-U. and Muser, M.: *Seat Component to Prevent Whiplash Injury*. Patent EP 02 405 537.8, European Patent Office, 2002.
61. Wilson, R.A.: A review of vehicle impact testing: How it began and what is being done. SAE Paper no700403, SAE Transactions, SAE International, Warrendale, PA, 1970.
62. Tidbury, G.H.: Vehicle structural analysis: A survey. *Int. J. Vehicle Design* 1, (1980), .165–172.
63. Kamal, M.M. and Wolf, J.A. (Eds.): *Modern Automotive Structural Analysis*. Van Nostrand Reinhold Company, New York, 1982.
64. McHenry, R.R.: Analysis of the dynamics of automobile passenger restraint systems. *Proceedings of the 7th Stapp Car Crash Conference*, Playa del Rey, CA, November 11–13, 1963, pp. 207–249.
65. Lobdell, T.E.: Impact response of the human thorax. In *Human Impact Response: Measurement and Simulation*. Plenum Press, New York, 1973, pp. 201–245.
66. Nikravesh, P.E., Chung, I.S., and Benedict, R.L.: Plastic hinge approach to vehicle crash simulation. *Comp. Struct*. 16(1–4), 1983, pp. 395–400.
67. Ambrósio, J., Pereira, M., and Dias, J.: Distributed and discrete nonlinear deformations in multibody dynamics. *Nonlinear Dyn*. 10(4), (1996), 359–379.
68. Milho, J., Ambrósio, J., and Pereira, M.: Design of train crash experimental tests by optimization procedures. *Int. J. of Crashworthiness* 9(5), (2004), 483–493.
69. Robbins, D.H., Bowman, B.M., and Bennett, R.O.: The MVMA two-dimensional crash victim simulation. *Proceedings of the 18th Stapp Car Crash Conference*. SAE International, Warrendale, PA, 1974, pp. 657–678.
70. Fleck, J., Butler, F., and Vogel, S.: *An Improved Three-Dimensional Computer Simulation of Motor Vehicle Crash Victims*, Tech. Report no ZQ-5180-L-1, Calspan Corporation, Buffalo, New York. 1974.
71. Laananen, D.H., Bolukbasi, A.O., and Coltman, J.W.: Computer simulation of an aircraft seat and occupant in a crash environment: Volume 1, Technical Report DOT/FAA/CT-82/33-I, US Department of Transportation, Federal Aviation Administration, Atlantic City, NJ, 1983.
72. Thompson, J.E.: Vehicle crush prediction using finite element techniques, paper no 730157. SAE Transactions, SAE International, Warrandale, PA, 1973.
73. Belytschko, T. and Hsieh, B.J.: Nonlinear transient finite element analysis with convected coordinates. *Int. J. Numer. Meth Eng*., 7, (1973), 255–271.
74. Hallquist, J.O.: *Preliminary Users Manual for DYNA3D and DYNAP (Nonlinear Dynamic Analysis of Solids in Three Dimensions)*. Report UCID-17268. Lawrence Livermore National Laboratory, Berkley, CA, 1976.
75. Belytschko, T. and Kenedy, J. M.: *WHAMS-3D, An Explicit 3D Finite Element Program*. KBS2 Inc., Willow Springs, IL, 1986.
76. Nikravesh, P.: *Computer-Aided Analysis of Mechanical Systems*, Prentice Hall, Englewood-Cliffs, NJ, 1988.
77. Verissimo, P. and Ambrósio, J.: Improved bushing models for vehicle dynamics, In C.A. Mota Soares et al. (eds.), *Proceedings of III European Conference on Computational Mechanics: Solids, Structures and Coupled Problems in Engineering* (on CD), Lisbon, Portugal, June 5–8, 2006, pp. 1–20.
78. Lankarani, H.M., Ma, D., and Menon, R.: Impact dynamics of multibody mechanical systems and application to crash responses of aircraft occupant/structure, In: M. Pereira and J. Ambrósio (eds.) *Computational Dynamics in Multibody Systems*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995, pp. 239–265.
79. Wu, S., Yang, S., and Haug, E.: Dynamics of mechanical systems with coulomb friction, stiction, impact and constraint addition-deletion, Technical Report 84-19, Center for Computer Aided Design, University of Iowa, University Heights, IA, 1984.
80. Winmer, A.: Einfluss der lelastungsgeschwindigkeit auf das festigkeitsund verformungsverhalten am beispiel von kraftfarhzeugen. *ATZ*. 77(10), (1977), 281–286.
81. Sousa, L., Veríssimo, P., and Ambrósio, J.: *Generic Multibody Car Models*, Technical Report IST -APROSYS-SP7, Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal. 2006.
82. Puppini, R., Diez, M., Avalle, M., Ciglaric, I., and Feist, F.: Generic car (FE) models for categories super minis, small family cars, large family executive cars, MPV and heavy vehicle, Technical Report APROSYS AP-SP7-0029-A, 2005.

83. ABAQUS Analysis User's Manual. Hibbitt, Karlsson and Sorensen, Inc., Pawtucket, Richmond, CA, 2004.
84. Kang, W.J., Cho, S.S., Huh, H., and Chung, D.T.: Modified Johnson-Cook model for vehicle body crashworthiness simulation. *Int. J. Vehicle Design*. 21(4/5), (1999), 424–435.
85. J. Dias and J. Ambrósio: APROSYS-SP6: Actuator requirements definition multibody simulations. Technical Report AP-SP61-D614, 2005.
86. Lim, J.M., Choi, J.W., and Park, G.-J.: Automobile side impact modeling using ATB software. *Int. J. Crashworthiness* 2(3), (1997), 287–297.
87. Zhu, J.Y., Cavanaugh, J.M., and King, A.I.: *Pelvic Biomechanical Response and Padding Benefits in Side Impact Based on a Cadaveric Test Series*. SAE Paper no 933128, Warrendale, PA, 1993, 223–233.

Automotive Engineering

Featuring contributions from leading experts, the ***Road and Off-Road Vehicle System Dynamics Handbook*** provides comprehensive, authoritative coverage of all the major issues involved in road vehicle dynamic behavior. While the focus is on automobiles, this book also highlights motorcycles, heavy commercial vehicles, and off-road vehicles.

The authors of the individual chapters, either from the automotive industry or universities, address basic issues, but also include references to significant papers for further reading. Thus the handbook is devoted both to the beginner, wishing to acquire basic knowledge on a specific topic, and to the experienced engineer or scientist, wishing to have advanced information on a particular subject. It can also be used as a textbook for master courses at universities.

The handbook begins with a short history of road and off-road vehicle dynamics followed by detailed, state-of-the-art chapters on modeling, analysis and optimization in vehicle system dynamics, vehicle concepts and aerodynamics, pneumatic tires and contact wheel-road/off-road, modeling of vehicle subsystems, vehicle dynamics and active safety, man-vehicle interaction, intelligent vehicle systems, and road accident reconstruction and passive safety.

- Provides extensive coverage of modeling, simulation, and analysis techniques
- Surveys all vehicle subsystems from a vehicle dynamics point of view
- Focuses on pneumatic tires and contact wheel-road/off-road
- Discusses intelligent vehicle systems technologies and active safety
- Considers safety factors and accident reconstruction procedures
- Includes chapters written by leading experts from all over the world

This text provides an applicable source of information for all people interested in a deeper understanding of road vehicle dynamics and related problems.

3322

ISBN: 978-0-8493-3322-4
90000
9 780849 333224

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK